M. H. Birnbaum[1]

Problems with So-Called "Direct" Scaling

REFERENCE: Birnbaum, M. H., **"Problems with So-Called "Direct" Scaling,"** *Selected Sensory Methods: Problems and Approaches to Hedonics, ASTM STP 773,* J. T. Kuznicki, R. A. Johnson, and A. F. Rutkiewic, Eds., American Society for Testing and Materials, 1982, pp. 34–48.

ABSTRACT: It was once thought that magnitude estimation was a better method for the measurement of subjective value than the method of category rating. However, recent research shows that magnitude estimations do not have advantages over category ratings and are in fact no more than ordinal scales of subjective value. When subjects are asked to judge "differences" and "ratios" of subjective value, it appears that subjects compute subjective intervals for both tasks. Magnitude estimations of "ratios" appear to be an approximate exponential function of subjective differences, and category ratings appear to be an approximate linear function of subjective differences, though the exact form of these functions depends on the stimulus and response distributions. Scales of subjective value that are derived from the subtractive model of within-mode stimulus comparison appear to be largely independent of the stimulus spacing and the response procedure. Measurements of subjective value should be derived from a theory of empirical relationships, and should show generality across contexts and across empirical domains. So-called "direct" scales have not been successful in predicting empirical relationships, whereas subtractive model scale values have shown promise.

KEY WORDS: scaling, subjective measurement, category ratings, magnitude estimation, psychophysical law, ratio, difference judgment

Measurement psychologists are a strange lot. If you meet one, do not say, "How do you do?"—the question will be taken seriously. "How are you?" is, of course, the most popular psychological measurement question asked today. Fortunately, few people take it literally. Does it make sense to discuss private experiences of pleasure and annoyance (not to mention pain)?

To avoid philosophical objections to introspective reports, psychologists of the 1950s operationally defined psychological values in terms of the

[1]Associate professor of psychology, Department of Psychology, University of Psychology, University of Illinois, Champaign, Ill. 61820.

numerical responses that observers give in so-called "direct" scaling studies. This use of operational definition immediately got them into trouble, and it still does. A psychologist took his girlfriend to see Bo Derek in the movie, "*10*," and was himself rated by his girlfriend on a 10-point scale. She said he was a "5." "Well, how would I rate on a 100-point scale?" he said. She responded "33." He felt better. He had to! By definition, he caused her to like him better by using a 100-point scale instead of a 10-point scale. After all, 33 is greater than 5.

This story illustrates that one should distinguish numerical responses from subjective values, unless we wish to conclude that her liking was changed. A theory of judgment is needed to permit a meaningful comparison of numbers from different scales in order to compare subjective sensations.

Outline of Judgment

In order to clarify the discussion, it will be useful to separate judgment into two components. The psychophysical function is defined as follows

$$s_i = H(\Phi_i) \tag{1}$$

where

Φ_i = physical value of stimulus i,
s_i = subjective value, and
H = psychophysical function (sometimes called a "law" by those of a determined nature).

The judgment function expresses the relationship between subjective value and overt numerical response as follows

$$R_i = J(s_i) \tag{2}$$

where

R_i = the overt response,
s_i = subjective value, and
J = the judgment function.

"Direct" Scaling

Numerous procedures for obtaining "direct" scales of sensation have been proposed. In the method of category rating, the observer is presented with two extreme stimuli, given two corresponding responses, and told to rate the others in between. For example, if Phyllis Diller is 1 and Bo Derek is 10, how beautiful are X, Y, and Z?

In the method of magnitude estimation, the observer is instructed to respond with numbers that reflect the "ratios" of subjective value.[2] Sometimes, a standard and associated modulus are given. For example, if Bo Derek is 10, how beautiful are X, Y, and Z?

One might think that these two sets of numbers, category ratings (C) and magnitude estimations (M) would be linearly related. Instead, M is typically exponentially related to C [*1–4*].[3] Thus, if we write the following equations

$$M_i = J_M(s_i) \tag{3}$$

$$C_i = J_C(s_i) \tag{4}$$

the relationship between magnitude estimations and ratings can be written

$$M_i = J_M J_C^{-1}(C_i) \tag{5}$$

Because M and C are nonlinearly related, it follows that J_M and J_C cannot both be linear.

The Great Debate

Because two equally plausible operational definitions of sensation failed to agree, something had to be done. But there was no reason to prefer one procedure over the other. Therefore, a lot of nonsense has been written arguing that one procedure is better than another, morally superior, etc. A good number of puns can be found in the literature. For example, just because subjects are asked to judge "ratios" does not mean that a ratio model would fit their data, or that the resulting numbers are unique to a ratio scale. Yet, it has been argued that magnitude estimations of "ratios" yield a ratio scale [*3,4*]. Similarly, category scales have been called "confusion" scales in which truth cannot arise from confusion. Such puns should have no place in scientific debates, at least not when they are to be taken seriously.

The following arguments were proposed and refuted:

1. If we define M as the true measure of sensation, it follows that C is biased and invalid. Problem: if we define C as the true measure of sensation, then M is biased and invalid.

2. Suppose it is assumed that a power function describes H; therefore, J_M must be a similarity transformation. How do we know H is a power function?

[2]Quote marks are used to designate "ratio" and "difference" tasks or judgments obtained with these tasks. Quotations are not used for actual ratios and differences or theoretical statements.

[3]The italic numbers in brackets refer to list of references appended to this paper.

Because $R = J[H(\Phi)] = a\ \Phi^b$, and we assume J is a similarity transformation. Problem: if we assume Fechner's law, J is exponential.

3. Subjects cannot judge differences. How do we know? Because C is nonlinearly related to M. Problem: From this evidence it could be equally well argued that subjects cannot judge ratios.

4. Subjects have greater freedom in M experiment than in C. Problem: Freedom is certainly a good thing in politics, but only the pun makes it seem reasonable in psychophysics. It could be argued that the experimenter has less control.

Unless something is known or assumed about H, nothing can be inferred concerning J. Therefore, the above arguments are not sound. The data consist of a graph of the composition of $J[H(\Phi)]$. Clearly, J cannot be defined or assumed to be linear if two "direct" measures are nonlinearly related.

Contextual Effects

The relationships between M, C, and Φ depend on a number of factors: stimulus range, response range, stimulus distribution, value of standard (if any), and so on. Not only are M and C not linearly related to each other, neither M nor C can be relied upon to be linearly related to itself. This point deserves amplification. Magnitude estimations ought to be independent of the numerical examples used in instructions. If people really judge the ratio of the subjective magnitudes of two stimuli, it should not matter whether the experimenter explains the task using an example of "three times" or "nine times."

However, the supposed "freedom" that the subject enjoys to choose any number is illusory. Evidence shows that the subjects use the range of responses suggested in the experiment. For example, Mellers and Birnbaum [*1*] asked subjects to judge "ratios" of the darknesses of dot patterns. If the largest example response in the instructions was "3," the largest mean judgment was 3.4. When the largest example was "9," the same two stimuli received a mean judgment of 8! Thus, the judged "ratio" of the same two stimuli can be drastically influenced by a change in instructions that should have no effect.

Mellers and Birnbaum [*1*] also varied the spacing of the stimuli. They found that the judged "ratio" of the darkness of a square with 40 dots to a square with 12 dots was about "5.5," if 8 of the 11 stimuli presented have fewer than 40 dots. However, if only 3 patterns had fewer than 40 dots, the same two stimuli (40/12) received a judgment of only about "3.5!" This same pair of stimuli received "ratio" judgments that varied from about 2 to 5.5 across four different contextual conditions that were supposed to have no effect on magnitude estimation. If one attempted to fit power functions to the stimulus-response relationship, these judgments would imply exponents ranging from 0.58 to 1.42!

Both C and M depend on the stimulus range and response range. Those who believe in power functions like to compute the ratio of log response range to log stimulus range (the exponent), and think this ratio represents something about perception. However, both of these variables (stimulus and response range) appear to be under the experimenter's control. Therefore, this index for magnitude estimation appears to be as arbitrary as calculating the same index for category ratings.

It is useful to consider the evidence that would have supported the "direct" scaling methods. Imagine a world in which magnitude estimation, for example, provided a "valid" measure of sensation on a ratio scale. If magnitude estimation were a "valid" scale, scales obtained with this procedure would be useful for predicting something. The first thing magnitude estimations of "ratios" should predict is other judgments of "ratios." If the ratio of A to B is 2, and the ratio of B to C is 3, then the ratio of A to C is 6. $(A/B)\cdot(B/C) = A/C$.

However, "direct" estimations of "ratios" obtained with different groups of subjects (the usual procedure) violate this prediction. For example, Mellers and Birnbaum [1] found that the judged "ratio" of A to B was 5, the "ratio" of B to C was 3, yielding a prediction of 15 for the "ratio" of A to C. Instead, the actual "ratio" judgment was only about 7.5. If "direct" judgments of "ratios" do not predict other "direct ratios," what could one possibly hope them to predict?

The situation is even worse for it turns out that magnitude estimations do not constitute even an ordinal scale between-subjects. Mellers and Birnbaum [1] found that the judged "ratio" of s_{40}/s_{12} is "5.5," another group of subjects judged the "ratio" of s_{60}/s_{12} to be "5." Therefore, a square with 40 dots is supposed to be darker than a square with 60 dots! However, within subjects (within contexts), judges always say the 60 dot pattern is darker than the 40 dot pattern.

People realize that these variables affect judgment. Unfortunately, some have reacted to contextual effects as "noise," "friction," or minor nuisances that divert one from the big picture. They might say I am making a big mistake, or otherwise I would not be seeing these contextual effects. However, it is now becoming clear that the results of "direct" scaling studies depend almost entirely on these variables.

Cookbook Answers are Half-Baked

How do we deal with contextual effects? It has been argued that there is a "right" way to do psychophysics: do not use a standard, do not give any examples, use as big a stimulus range as possible, space the stimuli in geometric steps, use the proper response range for your stimuli, encourage the subjects to use nonintegers like e or π, and so on. This advice is usually without empirical or theoretical foundation, and it is often contradictory. Occasionally, it is based on an untested armchair theory.

For example, the advice to use no standards in magnitude estimation is sometimes offered as a solution to the fact that the "direct" scales for different standards are not linearly related, which would violate the ratio model. By using no standards, it becomes impossible to test the ratio model. However, if $R = J(s)$, and if J depends nonlinearly on the standard, how do we know that when we use no explicit standards the "average" J will be linear?

Consider the following debate between a "direct" scaler and a skeptic concerning context effects.

DIRECT: "Everyone knows that subjects have trouble making these judgments. There are end-effects, modulus effects, number preferences, and so on. You need to know how to do these studies properly, otherwise you get biased results."

SKEPTIC: "How do you know the results are 'biased'?"

DIRECT: "Because they are nonlinearly related to what I get."

SKEPTIC: "How do you know your results are not biased?"

DIRECT: "Because I do the experiment properly, and I don't get context effects."

SKEPTIC: "You don't get an effect of the standard?"

DIRECT: "That is right, I don't use a standard."

SKEPTIC: "You don't get an effect of stimulus spacing and range?"

DIRECT: "Of course not. I've never seen effects of that kind because I control them. I always use the same range and spacing."

SKEPTIC: "How do you know you have the right stimulus range and spacing?"

DIRECT: "Because the books on scaling tell us to use the largest comfortable range and a geometric spacing. The books know the right answer."

SKEPTIC: "But different books say different things. How do you know which books to believe?"

DIRECT: "I only believe the right books."

SKEPTIC: "How do you know the books you read are right?"

DIRECT: "Because they don't get biased results."

The attitude seems to be as follows: do not do experiments that demonstrate that the scale values are wrong. Instead, design studies to avoid evidence against the accepted theories. If Aristotle believes that velocity of a falling object increases with mass, and the evidence for lead and aluminum balls shows a "bias," we should use only objects that conform to the theory; for example, feathers and pennies.

If we consider that magnitude estimations and category ratings are given by Eqs 3 and 4, and that the J_M and J_C functions vary nonlinearly as a function of such contextual manipulations as value of standard, modulus, stimulus range and spacing, and response range, it follows that unless we understand exactly how these factors affect J, we cannot decide *a priori* on the "right" procedures that guarantee J is linear. By allowing each subject to choose his or her own standard and modulus, we produce an "average" J function that seems unlikely to be linear. By selecting any given procedure *a priori*, we advocate a methodology based on an empirical theory that cannot be tested within the methodology advocated.

Measurement and Models

In recent years, psychologists have rejected the view that overt responses can be operationally defined as subjective values. Instead, the approach of conjoint and functional measurement has been to regard measurements as parameters of a theory of a nontrivial data array [5,6]. The uniqueness of measurements then depends on the family of transformations possible, under which the fit (of the theory to the rank order of the data) is invariant.

It used to be said, "measurement consists of assigning numbers to objects (or attributes thereof) according to rules." Psychological measurement then seemed to consist of using a subject to assign numbers according to rules posed by the instructions. This view has now given way to the idea that measurement consists of the construction of homomorphisms between relational structures, where one structure represents an empirical domain with experimental operations and empirical relations, and the other structure is mathematical. In this view, the empirical operations and relations require operational definitions, but the existence of the homomorphism to a mathematical measurement structure and the values on the scale are not matters of definition.

According to the modern measurement viewpoint, category ratings and magnitude estimations in "direct" scaling studies are no more than ordinal scales. Any monotone transformation of the data preserves the ordinal information in a typical "direct" scaling study.

To claim that magnitude estimations of "ratios" are supposed to represent subjective ratios, certain predictions should be verified. Suppose

$$R_{ij} = J[s_j/s_i] \tag{6}$$

if J is a power function, then it follows that

$$R_{ij} = R_{ik}R_{kj} \tag{7}$$

If sufficient data are gathered, independently manipulating s_j and s_i, it is possible to derive the values of s from the data, using Eq 6. Because any power transformation of s will reproduce the rank order of the data equally well, the ratio model yields scales that are unique to a power function (a log-interval scale, not a ratio scale, as the pun with the task would suggest).

In the measurement framework, bigger experiments are required to derive scale values. These experiments not only allow one to derive scales, but they also permit tests of the theory used to scale the stimuli. Space does not permit a complete discussion of measurement theory and the techniques for decomposing the functions. See Krantz, Luce, Suppes, and Tversky [*5*] for a discussion of conjoint measurement, and Anderson [*6*] for a brief review of functional measurement. The principle of stimulus scale convergence and the scale-free test, additional tools of model analysis, are presented by Birnbaum [*2, 7*]. These references describe how experiments with greater ordinal constraint can be analyzed to address issues that cannot be resolved in the "direct" scaling framework.

Experimental Test

Ten years ago, Clairice Veit and I decided to put category ratings and magnitude estimations to the test in their own domains, "differences" and "ratios." We thought that the data obtained with one procedure or the other would need monotonic rescaling in order to remove nonlinear response bias in J. We expected to be able to derive a single scale that would reproduce both "ratio" and "difference" matrixes. We used factorial designs in which tests of the models are nontrivial. To our surprise, both models fit their numerical data! However, the scale values derived from the two models did not agree. When both sets of data were separately transformed to the same model, then the scale values agreed. Suddenly, we had good evidence for a previous conjecture of Torgerson [*8*] that subjects perceive a single relationship between two stimuli. Since publication of that paper [*9*], the accumulated evidence has strengthened belief in the proposition that subjects use the same operation whether instructed to judge "ratios" or "differences."

"Ratios" and "Differences": One Operation

It turns out that judgments of "ratios" (R) and "differences" (D) are monotonically related

$$R_{ij} = M(D_{ij}) \tag{8}$$

where

M = a monotone function.

If subjects actually use both ratio and difference operations as instructed, one would not expect these two judgments to be monotonically related. For example

$$7 - 5 > 2 - 1, \quad \text{but} \tag{9}$$

$$7/5 < 2/1. \tag{10}$$

Thus, if subjects actually used two operations as instructed, we would not expect "ratios" and "differences" to be monotonically related. On the other hand, if subjects use the same operation despite instructions, we would expect them to be monotonically related. Birnbaum [*10*] has reviewed 9 experiments in which judgments of "ratios" and "differences" conform to the theory that subjects compare stimuli in the same way regardless of instructions. Five other experiments, reviewed by Birnbaum [*7*], gave similar conclusions. Schneider [*11*] has summarized further evidence showing that for a variety of continua (but not visual length), "ratios," and "differences" involve the same process.

Indeterminacy?

Torgerson [*8*] concluded that if subjects use only one operation, it would not be possible to discover empirically what operation was used. He argued that those investigators who favor ratio models and magnitude estimation could never convince those who favor subtractive models and category ratings and vice versa. The ordinal implications of the ratio model

$$R_{ij} = J(s_j/s_i)$$

and the subtractive theory of "ratio" judgments

$$R_{ij} = J^*(s_j^* - s_i^*)$$

are equivalent. Both theories would be refuted by the same evidence, for logarithmic transformation converts a ratio to a difference, and exponential transformation converts a difference to a ratio. According to Torgerson [*8*], there would be no way to resolve the debate, because the debate is empirically meaningless, a mere dispute over definitions.

A Solution to the Controversy

A series of experiments [*2,7,12,13*] has recently provided the theoretical and empirical constraint necessary to distinguish the two theories.

Birnbaum [2, 7] has shown that theories make differential ordinal predictions for experiments in which observers are asked to judge relations between stimulus pairs. For example, consider four stimuli, A, B, C, and D. Subjects could be asked to judge the "ratio of their differences," $(A - B)/(C - D)$; the "difference between their ratios," $(A/B) - (C/D)$; the "difference of differences," $(A - B) - (C - D)$; or the "ratio of ratios," $(A/B)/(C/D)$. Without knowing the subjective values of the stimuli and the exact form of the J function, it is possible to distinguish the theories. To illustrate, assume $A > B > C > D$. If a ratio of differences model applies to "ratios of differences," it follows

$$\frac{A - D}{A - C} < \frac{B - D}{B - C} \tag{11}$$

For example

$$\frac{7 - 1}{7 - 3} < \frac{4 - 1}{4 - 3}$$

when A, B, C, and D, 7, 4, 3, and 1, respectively. On the other hand, the difference of ratios model implies

$$A/D - A/C > B/D - B/C \tag{12}$$

For example

$$7/1 - 7/3 > 4/1 - 4/3$$

A difference of differences model or ratio of ratios model implies both judgments should be equal

$$(A/D)/(A/C) = (B/D)/(B/C) \tag{13}$$

$$(A - D) - (A - C) = (B - D) - (B - C) \tag{14}$$

Therefore, it is possible to use the rank order of the responses to determine which model is appropriate for each set of data. In addition, the theories are further constrained by the principle of scale convergence—the same subjective values should operate in all of the tasks and models.

These experiments [2, 7, 12, 13] provide evidence consistent with the subtractive theory of stimulus comparison, yielding the following conclusions:

1. The subjective magnitudes of stimuli can be represented as points along a continuum with no well-defined zero point. The subjective magnitudes appear to be largely independent of task, instruction, or response procedure.

2. When subjects are instructed to judge either "ratios" (R) or "differences" (D), the subtractive model can be used to represent data for both tasks

$$R = J_R\,[A - B] \tag{15}$$

$$D = J_D\,[A - B] \tag{16}$$

3. The ratio of differences model describes "ratios of differences," (RD); but the difference of differences model can be applied to the data for the following three tasks: "differences of differences," (DD); "differences of ratios," (DR); and "ratios of ratios," (RR)

$$RD = J_{RD}\,[(A - B)/(C - D)] \tag{17}$$

$$RR = J_{RR}\,[(A - B) - (C - D)] \tag{18}$$

$$DD = J_{DD}\,[(A - B) - (C - D)] \tag{19}$$

$$DR = J_{DR}\,[(A - B) - (C - D)] \tag{20}$$

4. The output, or judgment functions depend on the stimulus and response distributions and the procedure for response. The typical J function for category ratings is approximately linear, and the typical J function for magnitude estimations is nearly exponential. Variations in the J function can be approximated by an extension of Parducci's range-frequency theory.

5. Contextual effects are large for category ratings, magnitude estimations, and cross-modal judgments. However, the contextual effects are lawful enough to permit measurement of subjective value.

6. Scale values derived from the subtractive model applied to "ratio" and "difference" judgments are largely independent of the stimulus distribution.

Problems for the Ratio Theory

It seems reasonable to ask: "How can one save the ratio theory? What modifications of the theory would be necessary in order for it to be reasonable?" Veit [*13*], Birnbaum [*2, 7, 10*], Eisler [*14*], and Rule and Curtis [*15*] have discussed this question. In order to conclude that a ratio model is appropriate for "ratio" judgments, it appears one would have to deal with the following problems:

1. Birnbaum and Mellers [*16*] asked subjects to judge "ratios" and "differences" of easterliness and westerliness of U.S. cities. For example, what is the ratio of the easterliness of Philadelphia to that of Denver? Mental maps

of the United States based on ratio theory were found to be different depending on whether the subject is judging "easterliness" or "westerliness." Neither map based on ratio theory resembled the actual U.S. map. On the other hand, the subtractive theory implies only one mental map of the United States, one that resembles the actual U.S. map [*16*].

2. Ratio theory implies that subjective pitch is not linearly related to the musical scale. The subtractive model implies a scale that agrees with the musical scale [*17*].

3. Ratio theory implies that the psychophysical values of numbers are a positively accelerated function of actual number. The subtractive theory yields a negatively accelerated psychophysical function for number that agrees with previous results including the scale derived from range-frequency theory [*18,19,20*].

4. Ratio theory implies that "ratios of differences" (*RD*) are not calculated according to a ratio model, but instead according to the following exponential model

$$RD = (a/b)^{\frac{1}{c^* - d^*}} \tag{21}$$

This model assumes that the "difference" instruction sometimes corresponds to a ratio and sometimes to the log of a ratio, and that the "ratio" instruction sometimes corresponds to a ratio and sometimes to an exponential. Furthermore, two scales are required [*2,13*]. The subtractive theory represents "ratios of differences," by a ratio of differences model, using one scale of subjective value.

5. Eisler [*14*] suggested saving the ratio theory by introducing the assumption that internal transformations change the scale values depending on the instructions. Because this transformation theory makes an erroneous prediction for "difference of difference" and "difference of ratio" experiments, Eisler [*14*] was forced to add a *post hoc* postscript to the theory that subjects "reinterpret" one task. With the "reinterpretation" of the theory, it is ordinally equivalent to subtractive theory. Eisler's [*14*] prediction for variances was not supported by data [*7*].

6. Rule and Curtis [*15*] suggested that subjects may use both ratio and difference operations for corresponding tasks. To save the ratio theory for "ratio" judgments with an extension of the theory of Rule and Curtis, Birnbaum [*10*] found it would be necessary to assume that the output function for magnitude estimations of "ratios" would have to exceed 3 and often to exceed 4. Thus, when a subject says one stimulus is "16 times" as intense as another, the ratio theory would conclude it was actually only twice as intense. In the theory of Rule and Curtis, the output exponent is supposed to be the reciprocal of the input exponent for number. They assume the output expo-

nent is approximately 1.5. Therefore, values above 3.0 are clearly unacceptable to their theory of magnitude estimation.

Problems with "Direct" Scales

We can summarize the problems with so-called "direct" scales as follows:

1. Theoretically a "direct" scale represents a composition of functions; for example magnitude estimations can be represented as follows

$$M = J[s]$$

In a within-subject design, if $M_i > M_j$, it is usually the case that $C_i > C_j$ and $s_i > s_j$, where s has been derived from the subtractive model of "ratio" and "difference" judgments. Therefore, M and C should be regarded as ordinal scales. There is no reason to assume J is linear for any particular procedure advocated.

2. Context effects due to variation in the standard, modulus, stimulus spacing, example responses, and so on, can be represented by changes in the J function.

3. "Direct ratio scales" fail to predict other "direct" scales of equal face validity.

4. "Direct ratio scales" fail to predict "ratio" judgments.

5. "Direct" scales have not had success in predicting judgments of "differences," or other psychophysical relationships.

Advantages of Measurement-Derived Scales

Scale values derived from the subtractive model of "differences" and "ratios" show the following advantages:

(*a*) Differences between scale values can be used to reproduce the rank order of "differences" judgments and are theoretically unique to an interval scale.

(*b*) Theoretically, the H and J functions can be decomposed in multifactor research.

(*c*) Scale values derived from the subtractive model do not appear to depend on stimulus spacing, even though "direct" category or magnitude scales based on the same stimuli do depend on these variables.

(*d*) Scale values derived from the subtractive theory of stimulus comparison are consistent with scale values derived from range-frequency theory applied to category ratings of numerical magnitude, although the J functions for category ratings are lawfully nonlinear as a function of the stimulus spacing [*18,19*].

(*e*) Scale values derived from an additive model of combination agree with scale values used to predict single ratings of components, although *J* functions are lawfully nonlinear depending on the stimulus distribution [*1*].

Final Word

What is the point of measuring subjective value anyway? To what purposes will the scales be put? Scale values are often used to represent the utilities of products or the consequences of actions. When factors such as cost increase as utility increases, optimum decisions depend on a proper measurement of utility. Under these conditions, we need to have more than an ordinal scale of subjective value.

People used to argue that magnitude estimations provide a ratio scale of sensation, that they are the best way to measure sensation, and that the psychophysical law is a power function. The arguments for these contentions do not appear convincing. Instead, it appears more tractable to theorize that magnitude estimations are an ordinal scale of subjective value. Therefore, "direct" judgments do not reveal subjective value directly, unless we wish to return once more to the operational definition that gives us philosophical problems and renders the debate empty.

References

[*1*] Mellers, B. A. and Birnbaum, M. H., "Loci of Contextual Effects in Judgment," Illinois Judgment Laboratory Report, No. 1, 1980.

[*2*] Birnbaum, M. H., "Differences and Ratios in Psychological Measurement," *Cognitive Theory*, Vol. 3, N. J. Castellan and F. Restle, Eds., Lawrence Erlbaum Associates, Hillsdale, N. J., 1978.

[*3*] Stevens, S. S., *Psychological Review*, Vol. 78, 1971, pp. 426-450.

[*4*] Stevens, S. S. and Galanter, E. H., *Journal of Experimental Psychology*, Vol. 54, 1957, pp. 334-411.

[*5*] Krantz, D. H., Luce, R. D., Suppes, D., and Tversky, A., *Foundations of Measurement*, Academic Press, New York, 1971.

[*6*] Anderson, N. H., *American Scientist*, Vol. 21, 1979, pp. 201-215.

[*7*] Birnbaum, M. H., "Controversies in Psychological Measurement," *Social Attitudes and Psychophysical Measurement*, B. Wegener, Ed., Lawrence Erlbaum Associates, Hillsdale, N. J., 1982.

[*8*] Torgerson, W. S., *Acta Psychologica*, Vol. 19, 1961, pp. 201-205.

[*9*] Birnbaum, M. H. and Veit, C. T., *Perception & Psychophysics*, Vol. 15, 1974, pp. 7-15.

[*10*] Birnbaum, M. H., *Journal of Experimental Psychology: General*, Vol. 109, 1980, pp. 304-319.

[*11*] Schneider, B., "The Nonmeteic Analysis of Difference Judgments in Social Psychophysics: Scale Validity and Dimensionality," *Social Attitudes and Psychophysical Measurement*, B. Wegener, Ed., Lawrence Erlbaum Associates, Hillsdale, N. J., 1982.

[*12*] Hagerty, M. and Birnbaum, M. H., *Perception & Psychophysics*, Vol. 24, 1978, pp. 121-129.

[*13*] Veit, C. T., *Journal of Experimental Psychology: General*, Vol. 107, 1978, pp. 81-107.

[*14*] Eisler, H. *Perception & Psychophysics*, Vol. 24, 1978, pp. 185-189.

[*15*] Rule, S. J. and Curtis, D. W., *Journal of Experimental Psychology: General*, Vol. 109, 1980, pp. 296-300.

[*16*] Birnbaum, M. H. and Mellers, B. A., *Perception & Psychophysics,* Vol. 23, 1978, pp. 403–408.
[*17*] Elmasian, R. and Birnbaum, M. H., "A Harmonious Note on Pitch," Illinois Judgment Laboratory Report, No. 1, 1979.
[*18*] Birnbaum, M. H.; *Perception & Psychophysics,* Vol. 15, 1974, pp. 89–96.
[*19*] Rose, B. J. and Birnbaum, M. H., *Perception & Psychophysics,* Vol. 18, 1975, pp. 194–200.
[*20*] Parducci, A., "Contextual Effects: A Range-Frequency Analysis," *Handbook of Perception,* Vol. 2, E. C. Carterette and M. P. Friedman, Eds., Academic Press, New York, 1974.