On the ability to estimate differences: A note on Birnbaum's subtractive model

HANNES EISLER

University of Stockholm, S-113 85 Stockholm, Sweden

In a number of papers (e.g., Birnbaum & Elmasian, 1977), Birnbaum and his co-workers have argued, in agreement with Torgerson (Note 1, 1961), that subjects in psychophysical experiments are unable to distinguish between ratios and differences. They claim that the formation of ratios and the formation of differences involve the same operation, carried out on the same scale, and they attribute the observed nonlinear relation between "ratio" and "difference" scales to different judgment functions. (The quotation marks around "ratio" and "difference" indicate that these operations refer to the instruction, irrespective of what the subject does.) Since this claim cannot be substantiated (on the ordinal level) with unifactor stimulus designs, Birnbaum and his co-workers resorted to two-stimulus designs, with which they were able to confirm the notion of *one* operation on *one* scale (Birnbaum & Elmasian, 1977; Birnbaum & Veit, 1974; Rose & Birnbaum, 1975). However, a two-stimulus design does not enable one to determine by which of the two possible arithmetical operations, division or subtraction, the subject's comparison operation should be represented. Subsequent development of the measurement theory (Birnbaum, 1978) seemed to show that with a four-stimulus design one could arrive at a decision as to the "appropriate" scale and the "correct" nature of the ratio-difference operation. Birnbaum (1978) is inclined to conclude that it is subtraction that yields the "correct" scale values. In this note, I shall dispute his conclusions—both that a distinction can be made with the four-stimulus design and that the subtractive model with the appurtenant scale is the "correct" interpretation.

Forget for a moment about psychophysics and imagine that you ask somebody (with veridical sensation) to judge the ratio between a 600-g and a 300-g weight. The answer will be "two to one," or "twice as heavy," or "200 percent." When asked to judge the difference, on the other hand, the reply will have to be "300 grams" or "about a third of a kilo." Unlike a judged ratio, a judged difference carries no meaning unless one specifies a denomina-

tion or unit. This contention applies only to quantitative judgments of differences; if differences only are to be equated, as in equisection or category rating, there is no need to specify a unit. We can conclude that quantitative estimations of differences necessarily require, over and above the difference operation, a ratio operation, viz, division by the prescribed unit. In the experiments described by Birnbaum (1978), the subject had to estimate both ratios and differences relative to a standard. While this is supererogatory regarding ratios, it is bitter necessity for differences.

Let us, in agreement with Birnbaum (1978), make the following fundamental assumptions: (1) independence, i.e., the scale value of a stimulus is independent of the stimuli with which it is compared, and (2) scale convergence, i.e., the scale value of a stimulus is independent of tasks. Let us furthermore assume that the subject is capable of performing *only one operation*, which we denote by $\ominus$.

Judgment Functions and Transformations

In order to account for the observed differences between ratio and difference scales, Birnbaum (1978) assumes a number of "judgment functions" which transform the outcome of the sometimes rather complex comparison operation(s) into the overt numerical response. Disregarding secondary effects, there are, however, only two such judgment functions; one of them is linear and, for our purpose, can be considered to be the identity function and accordingly omitted. Let us denote the other by T' . What I want to propose instead is likewise two transformations, of which one, being the identity transformation, likewise can be omitted; the other will be denoted by T . But, unlike T' , T is not considered to be a judgment function. Rather, it is applied whenever the subject is required to deal with differences. Thus, a ratio of s_j to s_i is denoted by $s_j \oslash s_i$, and a difference between the same scale values by $T(s_j \ominus s_i)$. The rationale—be it mine or the subject's—is to give a meaningless unit-free difference a tractable inner representation. With one- and two-stimulus designs, one cannot decide whether the transformations may be considered judgment functions or not. As we shall see below, with four-stimulus designs, the choice between the two ways of introducing the necessary transformations is crucial. As an example, let us analyze a "ratio of differences." The formation of such a judgment would have to proceed in two steps: first the formation of (two) "differences," then the formation of the "ratio" of these "differences." Since the "differences" have to be formed before the "ratio," T is applied to each "difference." A judgment function as used by Birnbaum (1978)

This investigation was supported by the former Swedish Council for Social Science Research. Requests for reprints should be sent to Hannes Eisler, Department of Psychology, University of Stockholm, Box 6706, S-113 85 Stockholm, Sweden.

would apply to the end result, i.e., to the inner representation of the "ratio of differences."

Measurement Theories

Birnbaum's (1978) measurement theory encompasses "ratios" and "differences" of four scale values in different combinations, besides the two two-stimulus combinations. Table 1 shows, for all six (combinations of) operations, how the present model and that of Birnbaum (1978) are to be represented. The present model is shown with operation and transformation both unspecified and specified. The specification shows that T is the logarithmic transformation and $\otimes$ division throughout. Birnbaum's (1978) model is given only with specified operations (subtraction and, for "ratio of differences," subtraction and division) and the pertinent judgment function (exponentiation). If only the rank order of the data is considered, however, Birnbaum's (1978) exponential transformation can be omitted.

Figure 1 shows which response patterns are to be expected from the present model, described in the manner employed by Birnbaum (1978, Figure 10) for an imaginary subject who uses subtraction and division correctly in the arithmetic sense on the same scale values. As can be seen by replacing $\log s$ by s^* , Birnbaum's (1978) and the present model make the same predictions as to the response patterns for "ratio of ratios," "ratio of differences," and "difference of ratios," and these patterns agree with what Hagerty and Birnbaum (1978, quoted by Birnbaum, 1978) obtained. In Figure 1, the upper left panel should be compared with Figure 15, upper panels, in Birnbaum (1978), the lower left panel with Birnbaum's Figure 14, and the upper right panel with Birnbaum's Figure 17.

The Problem of "Difference of Differences"

The "difference of differences" instruction, marked by a question mark in Table 1 and Figure 1, requires a detailed analysis. From Birnbaum's (1978)

Figures 16 and 17, it can be seen that the subjects make virtually the same estimations for the "difference of ratios" and the "difference of differences" tasks. This finding is interpreted by Birnbaum to mean that there the subjects do not distinguish between ratios and differences, perhaps somewhat surprisingly, since a different instruction according to Birnbaum's (1978) model supposedly brings with it an at least slightly different judgment function. I am inclined to explain this finding as indicating, instead, that the subjects are unable to carry out the "difference of differences" instruction. This is easily realized in the framework of the present model. "Difference of differences" would require $T[T(\cdot)]$, implying $\log[\log(\cdot)]$. Since $\log(\cdot)$ may be zero or negative, the transformations cannot always be carried out. What does a subject do in such a situation? I have practically never heard of a subject's telling the experimenter that the task required is impossible. Rather, she reinterprets the task; for the case at hand, "difference of differences" is reinterpreted as "difference of ratios." Reinterpretation of a psychophysical task by a subject is not unusual. One instance is to be found in Eisler and Roskam (1977a), where subjects were required to estimate "commonality ratios" but estimated similarities instead. Other instances occur in time perception studies (Eisler, 1975). What has been said above is clearly insufficient to decide between the two interpretations of subjects' making the same estimations for the two instructions. Perhaps a clue may be found in the standard deviations of the estimates. If Birnbaum's (1978) subtractive model is correct, the "difference of differences" instruction should be more congruent with the subjects' operation and thus yield smaller SDs. If, on the other hand, the subjects resort to reinterpretation of the "difference of differences" instruction, these SDs should be greater, because less consistency in the subjects' judgments—both intra- and interindividually—is to be expected.

Reinterpretation of the "difference of differences"

Table 1
Operations and Transformations Underlying Combinations of "Ratio" and "Difference" Instructions
According to the Present Model and Birnbaum's (1978) Model

	Present Model		Birnbaum's (1978) Model	Two Operations and Two Scales with $s^* = \log s$
	Unspecified	Specified		
"Ratio"	$s_j \otimes s_i$	s_j/s_i	$\exp(s_j^* - s_i^*)$	s_j/s_i
"Difference"	$T(s_j \otimes s_i)$	$\log(s_j/s_i)$	$s_j^* - s_i^*$	$s_j^* - s_i^*$
"Ratio of ratios"	$(s_j \otimes s_i) \otimes (s_l \otimes s_k)$	$(s_j/s_i)/(s_l/s_k)$	$\exp[(s_j^* - s_i^*) - (s_l^* - s_k^*)]$	$(s_j/s_i)/(s_l/s_k)$
"Ratio of differences"	$T(s_j \otimes s_i) \otimes T(s_l \otimes s_k)$	$\log(s_j/s_i)/\log(s_l/s_k)$	$(s_j^* - s_i^*)/(s_l^* - s_k^*)$	$(s_j^* - s_i^*)/(s_l^* - s_k^*)$
"Difference of ratios"	$T[(s_j \otimes s_i) \otimes (s_l \otimes s_k)]$	$\log[(s_j/s_i)/(s_l/s_k)]$	$(s_j^* - s_i^*) - (s_l^* - s_k^*)$	$(s_j/s_i)^* - (s_l/s_k)^*$
"Difference of differences"	$T[T(s_j \otimes s_i) \otimes T(s_l \otimes s_k)]$	?	$(s_j^* - s_i^*) - (s_l^* - s_k^*)$	$(s_j^* - s_i^*) - (s_l^* - s_k^*)$

Note—Scale values are denoted by s and s^* , the "ratio-difference" operation by $\otimes$, and the "difference" transformation by T .

Figure 1. Theoretical response patterns for four four-stimuli comparison tasks according to the present model; see Table 1, column headed "specified." The scale values are set to the integers between 1 and 7. For the second term of the instructions, the computations are carried out for only three combinations, viz, scale values 7, 5, and 3 combined with scale value 1, so that only three curve families are given in each panel; compare with Figure 10 in Birnbaum (1978). The ordinates give theoretical responses; the abscissas (Figure A) are column sums of the response matrices, subject to linear transformations to obtain comparability. This implies that, for the upper left panel, they are the actual scale values, for the other panels, their logarithms.

instruction and Thurstonian scaling. There is an interesting case of a "difference of differences" instruction in Birnbaum and Elmasian (1977). They instructed their subjects to rate "the difference in loudness between the second tone and the first" on a 9-point category scale. Since a category rating instruction induces the "difference" operation (or transformation, or judgment function), this instruction implies "difference of differences." It is not surprising that the outcome agrees with what has been said before, but another point is worth noting. In this context, the term "difference" need not be conceived of as an (arithmetic) subtraction; it may also mean "unlikeness." This interpretation is supported by the experimenters' labeling of the categories: from "very, very much softer" via "equal" to "very, very much louder." This conception connects

with the two equivalent representations of Thurstone's Case V, i.e., " $s_j > s_i$ " can be represented equally well by " $s_j^* - s_i^* > 0$ " as by " $s_j^{**}/s_i^{**} > 1$," where $s^* = \log s^{**}$ (Eisler, 1965a). If we choose "unlikeness" as an interpretation of "difference" and " $s_j/s_i > 1$ " as the representation of " $s_j > s_i$," then we have a "difference of ratios" instruction. Perhaps it is a notion like this that lies behind the subjects' reinterpretation of "difference of differences" as "difference of ratios."

Model Comparisons Including Abandonment of Scale Convergence

Table 1 shows that the present model and Birnbaum's (1978) are equivalent regarding (1) independence, (2) scale convergence, and (3) need for just one transformation. The difference between the

two models lies in the different interpretations of their respective transformations and, first and foremost, in that Birnbaum's (1978) model requires two operations (subtraction always, supplemented by division for the "ratio of differences" instruction), whereas the present model requires only one, division (ratio formation). Thus, from the viewpoint of parsimony, the present model is to be preferred. This comparison holds if the metric properties of the data are taken into account. On the ordinal level, however, the two models become equivalent: the present one requires one transformation and one operation, Birnbaum's (1978) only two operations, since the exponential transformation is monotonic and thus can be disregarded.

On the other hand, it may be objected that the present model essentially operates with two scales. Whenever the subject is required to deal with differences, she may apply the logarithmic transformation to the arguments of the "difference" before carrying out the subtraction, a procedure that can be regarded as operating on another scale, viz, a logarithmic one. This way of conceiving all relations corresponds to what Birnbaum (1978) termed "two worlds." In a way, I find the "two worlds" view appealing. I am not sure that it is good scientific strategy to prefer parsimony to face validity, i.e., to interpreting the subjects' estimates according to instructions. Furthermore, if "difference" (category) judgments operate on a discrimination scale (see, e.g., Eisler, 1963)—with constant variability—and ratio judgments on a sensation scale—with increasing variability—there is no a priori reason why these two scales should be identical, but there are a priori reasons why they should be monotonically related. If this relation is logarithmic, we are back at the starting point: subtraction and division on logarithmically related scales can, according to the present model, be represented solely by division on just one scale. The last column of Table 1 gives the "two operations on two scales" representation for the six tasks. We see that the representations of the six tasks are consistent—with one exception. The "difference of ratios" is denoted by $(s_j/s_i)^* - (s_l/s_k)^*$. Rigorously, the * notation applies to scale values, not to their ratios. We also see that the representation in the present model of "difference of ratios," $\log[(s_j/s_i)/(s_l/s_k)]$, can be written $\log(s_j/s_i) - \log(s_l/s_k)$, i.e., just as a difference of ratios, where, however, the ratios are subject to the logarithmic transformation, which leads, in turn, to Birnbaum's difference-of-differences representation, $(\log s_j - \log s_i) - (\log s_l - \log s_k)$. To retain the "two worlds" view, one has to suppose that the ratios are formed on the sensation scale and the differences on a discrimination scale of ratios. I am not pleading the reasonableness of such a scale, merely pointing

out how measurement theory and substantive theory may be intertwined.

In this connection, the outcome of Veit's (1978) magnitude estimation of differences experiment deserves some comment. Unlike in Hagerty and Birnbaum's (1978) and her own "ratio of differences" experiments, the straight lines of the response pattern are divergent rather than parallel.¹ While Veit (1978) interprets this divergence in terms of a nonlinear judgment function, I attribute it to "imperfect logarithms" in the sense that a constant is added to the argument of the logarithm as is the case in category scales (Eisler, 1962a, 1962b).

Conclusion

The discussion about the appropriateness of the subtractive or the ratio model is highly reminiscent of the corresponding argumentation regarding the two equivalent representations of Thurstone's Case V (Eisler, 1965a). Bock and Jones (1968, pp. 65-66, p. 273) provided evidence in favor of the ratio representation. For direct scaling, Eisler (1965b) argued, likewise, for the ratio representation. Further evidence in favor of the ratio representation can be found in Eisler (1975), where magnitude estimation of durations yields roughly the same power function exponent as the one obtained from duration reproduction. Also, the attempt to quantitatively relate different ratio-level estimates of the same stimuli proved successful (Eisler & Roskam, 1977a, 1977b), thereby supporting the ratio model.

From the above it may be tentatively concluded that there is an alternative to Birnbaum's model, according to which combinations of "difference" and "ratio" instructions in psychophysical experiments can be represented by a single scale, the sensation scale, and a single operation, ratio formation. Whether this representation is preferable to Birnbaum's (1978) or to one with two operations on two scales is a question that requires further investigation.

REFERENCE NOTE

1. Torgerson, W. S. *Distances and ratios in psychophysical scaling* (Report No. 58 G-0014). Lexington, Mass: MIT, Lincoln Laboratory, October 1960.

REFERENCES

- BIRNBAUM, M. H. Differences and ratios in psychological measurement. In F. Restle & N. J. Castellan, Jr. (Eds.), *Cognitive theory* (Vol. 3). Hillsdale, N.J: Erlbaum, 1978.
- BIRNBAUM, M. H., & ELMASIAN, R. Loudness "ratios" and "differences" involve the same psychophysical operation. *Perception & Psychophysics*, 1977, 22, 383-391.
- BIRNBAUM, M. H., & VEIT, C. T. Scale convergence as a criterion for rescaling: Information integration with difference,

- ratio, and averaging tasks. *Perception & Psychophysics*, 1974, **15**, 7-15.
- BOCK, R. D., & JONES, L. V. *The measurement and prediction of judgment and choice*. San Francisco: Holden-Day, 1968.
- EISLER, H. Empirical test of a model relating magnitude and category scales. *Scandinavian Journal of Psychology*, 1962, **3**, 88-96. (a)
- EISLER, H. On the problem of category scales in psychophysics. *Scandinavian Journal of Psychology*, 1962, **3**, 81-87. (b)
- EISLER, H. Magnitude scales, category scales, and Fechnerian integration. *Psychological Review*, 1963, **70**, 243-253.
- EISLER, H. The connection between magnitude and discrimination scales and direct and indirect scaling methods. *Psychometrika*, 1965, **30**, 271-289. (a)
- EISLER, H. On psychophysics in general and the general psychophysical differential equation in particular. *Scandinavian Journal of Psychology*, 1965, **6**, 85-102. (b)
- EISLER, H. Subjective duration and psychophysics. *Psychological Review*, 1975, **82**, 429-450.
- EISLER, H., & ROSKAM, E. E. Multidimensional similarity: An experimental and theoretical comparison of vector, distance, and set theoretical models. I. Models and internal consistency of data. *Acta Psychologica*, 1977, **41**, 1-46. (a)
- EISLER, H., & ROSKAM, E. E. Multidimensional similarity: An experimental and theoretical comparison of vector, distance, and set theoretical models. II. Multidimensional analyses: The subjective space. *Acta Psychologica*, 1977, **41**, 335-363. (b)
- HAGERTY, M., & BIRNBAUM, M. H. Nonmetric tests of ratio vs. subtractive theories of stimulus comparison. *Perception & Psychophysics*, 1978, **24**, 121-129.
- ROSE, B. J., & BIRNBAUM, M. H. Judgments of differences and ratios of numerals. *Perception & Psychophysics*, 1975, **18**, 194-200.
- TORGERSON, W. S. Distances and ratios in psychological scaling. *Acta Psychologica*, 1961, **19**, 201-205.
- VEIT, C. T. Ratio and subtractive processes in psychophysical judgment. *Journal of Experimental Psychology: General*, 1978, **107**, 81-107.

NOTE

1. Veit (1978) claims a slight divergence in at least one set of her "ratio of differences" data. I found it hard to see.

(Received for publication April 4, 1978;
accepted April 11, 1978.)