Perception & Psychophysics
1983, 34 (4), 405-408

Reply to Zwislocki's views on "absolute" scaling

BARBARA A. MELLERS
University of California, Berkeley, California

Zwislocki and Goodman (1980) contend that if subjects are asked to assign numbers to stimuli without the restrictions of a standard or modulus, the resulting scale will be "absolute." Zwislocki and Goodman argue that if "absolute" scales exist, then subjects' responses in "absolute" scaling experiments should be invariant with respect to contextual manipulations known to affect category ratings and magnitude estimations. They obtained subjects' responses using the "absolute" scaling procedure in different conditions and decided that contextual effects were small enough for the data to be consistent with the absolute scale hypothesis.

Mellers (1983) argued that Zwislocki and Goodman (1980) provided no frame of reference for evaluating the magnitude of contextual effects in their experiments. Mellers (1983) compared the magnitudes of contextual effects for different scaling procedures, and found that the "absolute" scaling procedure was as susceptible to changes in the stimulus spacing as were category ratings and magnitude estimations. Because the same stimuli received different responses when presented in different stimulus distributions, Mellers argued that there was no reason to postulate an "absolute" scale. Zwislocki (1983) responded to Mellers (1983) by arguing that data either validate the "absolute" scale hypothesis or else they are "biased." Thus, no experiment in principle can refute the theory of "absolute" scales.

In this note, I would like to clarify my views on four issues and contrast them with the opinions expressed by Zwislocki (1983). These issues include: (1) definition of scale values, (2) scale types, (3) validation of scales, and (4) relative vs. absolute judgment.

Definition of Scale Values

Zwislocki (1983) defines an "absolute" scale as an invariant mapping from sensations to numbers. He states, "If the units of measurement cannot be designated arbitrarily by the experimenter, so that even a multiplicative transformation is excluded, the scale must become formally absolute" (p. 593). In other words, responses are equivalent to scale values when subjects are allowed to select their own standards and moduli in magnitude estimations.

An alternative definition is that scale values are parameters of a theory that can be refuted by empirical investigation. According to this definition, responses need not be taken at face value. In tasks such as those used by Zwislocki and Goodman (1980), responses are represented as the composition of two functions—the psychophysical function (H) and the judgment function (J) (Birnbaum, 1974a, 1978, 1982a, 1982b; Mellers & Birnbaum, 1982, 1983). Responses are assumed to be monotonically related to scale values as follows:

$$R_{ik} = J_k(s_i), \qquad (1)$$

where R_{ik} is the response to the i^{th} stimulus in the k^{th} context; J_k is a monotonic judgment function; s_i is the scale value that is related to its physical counterpart, Φ_i, by the psychophysical function, H, where $s_i = H(\Phi_i)$. The judgment function is theorized to depend on a variety of factors, including the task, the instructions, the response examples, the stimulus distribution, and other procedural details of the experiment. All of these factors are summarized by the word context, subscripted k.

Responses from tasks in which subjects combine or compare stimuli can be represented as the composition of three functions: the psychophysical function (H), the comparison function (C) relating subjective values to an overall impression, and the judgment function (J) (Anderson, 1974; Birnbaum, Parducci, & Gifford, 1971). For example, if subjects are asked to judge the "difference" in loudness between two tones, one representation of the data might be:

$$D_{ijk} = J_k[C(s_i,s_j)] = J_k(s_i - s_j), \qquad (2)$$

where D_{ijk} is the judged "difference" in loudness between stimulus i and j, J_k is the judgment function for differences, C is a subtractive model, and s_i and s_j are scale values of the i^{th} and j^{th} stimuli, respectively. Responses are not assumed to be described by the instructed operation; rather, a subtractive operation is empirically tested with data obtained in multifactor designs. It is logically possible to obtain data that would refute Equation 2, in which case the derived scale values would be without theoretical status.

Scale Types

Zwislocki's (1983) view of scale types is based on Stevens's (1951) scale classification. In this classification, scale types appear to be determined by the

The author wishes to thank Michael Birnbaum for many valuable comments and discussions and Elizabeth Hartka and Howard Terry for suggestions on earlier drafts. Requests for reprints should be sent to Barbara A. Mellers, Department of Psychology, University of California, Berkeley, California 94720.

instructions given to subjects. This view differs from the view that scale types are based on a theory of the empirical structure (Krantz, Luce, Suppes, & Tversky, 1971). According to this approach, scale types are defined by the uniqueness of the theory used to represent an empirical structure (data) with a numerical one (Birnbaum, 1978, 1980, 1982; Krantz et al., 1971). If it can be shown empirically that the rank order of a set of responses is equivalent to the rank order of the predictions from a model, than one can ask: How can the derived scale values (s_i in Equation 2) be transformed so that the model still reproduces the rank order of the data? The set of permissible transformations defines the scale types.

In single-stimulus tasks, such as those used by Zwislocki and Goodman (1980), there is insufficient constraint in the rank order of the responses to adequately distinguish among models and determine scale types beyond an ordinal scale (see Birnbaum, 1978, for a discussion). However, it is sometimes possible to determine scale types with tasks using multifactor designs. For example, if subjects are asked to judge "differences" between pairs of stimuli and the responses are ordinally consistent with a subtractive operation, then scales derived from the model would, in principle, be unique up to a linear transformation (Krantz et al., 1971).

Zwislocki's "absolute" scale hypothesis is unclear, because he has not proposed a model or operation (C in Equation 2) that would define measurements unique up to an "absolute" scale. Responses from single-stimulus tasks that were ordinally consistent with an "absolute" scale would be consistent with many other representations of the data. Thus, the uniqueness of the scale types has not been determined.

Validation of Scales

Zwislocki (1983) states that "bias in psychophysical experiments means that subjects do not follow accurately the experimenter's instructions" (p. 593). He argues that it is up to the experimenter to arrange the "proper" conditions so that subjects are able to follow the rules. In support of this idea, he points to an example in physics—the law of falling bodies—and suggests that the "proper" experimental conditions in which to test the theory are in a vacuum, for, under those conditions, a penny and a feather would fall to earth at the same rate. However, we do not necessarily reject Gallileo's law of falling bodies if we investigate it in the atmosphere and find that the penny and the feather fall at different rates. Such deviations have made us recognize that, in the atmosphere, the law is incomplete and that a theory of air friction is required.

Zwislocki (1983) attempts to validate "absolute" scales by examining whether the responses across two or more conditions are appropriately related. So-called "absolute" scales would require identical responses across conditions; "ratio" scales would require responses that were related by a multiplicative constant. Zwislocki (1983) states that one needs to find only a single set of conditions consistent with the scale to validate it, and "the discovery of circumstances that preclude the achievement of a scale does not constitute proof of its general lack of validity" (p. 593). Such a view leads one to wonder how Zwislocki would test the "absolute" scale hypothesis, as his criterion for validation makes the theory impossible to refute.

Instead, it can be argued that science progresses through our attempts to "invalidate" models, scales, and theories. One reason to prefer Gallileo's law to Aristotle's (objects fall at different rates) is because the theory of Gallileo (modified by Newton and others) can account for experiments both in the atmosphere and the vacuum, whereas Aristotle's fails in the vacuum. Thus, it is the *failure* of Aristotle's theory to predict the results, not the success of Gallileo's, that convinced many scientists to prefer Gallileo's theory.

Zwislocki's argument that a single success under the "proper" conditions validates the theory would not allow one to decide between two theories. It would only generate a dispute over the proper conditions. Clearly, nothing is learned if one of the theories under investigation is used to define the proper conditions for a test between two hypotheses.

Deviations from a theory tell us that this theory is incomplete or incorrect and often force us to rethink our position. For example, when data violate the ordinal properties of a model, we can reject that model and the scales derived from it. If data are ordinally consistent with a model, then additional constraints can be applied to investigate whether the scales have generality to other empirical structures.

One such constraint is the scale convergence criterion (Birnbaum, 1974b, 1978, 1982a, 1982b; Birnbaum & Veit, 1974; Mellers & Birnbaum, 1982, 1983). This criterion, which applies to data from different tasks with the same stimuli in the same context, asserts that scale values should be *independent* of the task. Scale values from different tasks should be related according to the uniqueness of the models. If they are, then one set of scales can be used for the different empirical structures. If they are not, then the scales are limited and/or one or both of the models should be rejected.

Relative vs. Absolute Judgment

A number of theorists have argued that judgments are relative rather than absolute (Birnbaum, 1974a, 1982a, 1982b; Helson, 1964; Johnson & Mullally, 1969; Parducci, 1963, 1974, 1982; Poulton, 1968; Restle & Greeno, 1970). Relativity theory (Birnbaum, 1974a, 1975, 1982a, 1982b; Mellers, 1982, 1983; Mellers & Birnbaum, 1982, 1983) asserts that the subject considers the relative position of a stimulus in its context when comparing or combining stimuli. This context is partially determined by the experimenter (i.e., the task, the instructions, the stimulus distribution, etc.) but is also influenced by a context outside the laboratory (i.e., the subject's past and present experiences).

Mellers (1983) investigated the effects of the laboratory context by manipulating the stimulus spacing and found that the responses obtained using the "absolute" scaling procedure and other procedures were generally consistent with a range-frequency interpretation of the judgment function, J_k (Parducci, 1963, 1974, 1982; Parducci & Perrett, 1971) and a single set of scale values, s_i. If invariance with respect to the context is desirable, then one might prefer the techniques described in Mellers and Birnbaum (1982, Experiment 2). In that experiment, subjects judged both "ratios" and "differences" of the dot patterns used by Mellers (1983). Scale values derived from a subtractive model for both tasks (s_i in Equation 2) were far more invariant across contexts than were measures of sensation based on the "absolute" scaling procedure (R in Equation 1).

Conclusion

Is there a reason to prefer the "absolute" scaling method? Mellers (1983) found that the procedure appeared to be as susceptible to contextual effects as are category ratings and magnitude estimations. In addition, the "absolute" scaling procedure increases response variability and reduces the power of any tests, including tests for contextual effects.

The "absolute" scaling hypothesis is unable to account for the results of Mellers (1983). Although Zwislocki (1983) claims that the data of Mellers are "biased," he offers no theory of the so-called "biases."[1] On the other hand, relativity theory can, in principle, account for the data of Zwislocki and Goodman (1980), as noted by Mellers (1983). Thus, it seems entirely possible that "absolute" scales are a consequence of relative judgment.

REFERENCES

Anderson, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2). San Francisco: Freeman, 1974.

Birnbaum, M. H. Using contextual effects to derive psychophysical scales. *Perception & Psychophysics*, 1974, **15**, 89-96. (a)

Birnbaum, M. H. The nonadditivity of personality impressions. *Journal of Experimental Psychology*, 1974, **102**, 543-561. (b)

Birnbaum, M. H. Expectancy and judgment. In F. Restle, R. Shiffrin, N. J. Castellan, H. Lindman, & D. Pisoni (Eds.), *Cognitive theory* (Vol. 1). Hillsdale, N.J: Erlbaum, 1975.

Birnbaum, M. H. Differences and ratios in psychophysical measurement. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3). Hillsdale, N.J: Erlbaum, 1978.

Birnbaum, M. H. A comparison of two theories of "ratio" and "difference" judgments. *Journal of Experimental Psychology: General*, 1980, **3**, 304-319.

Birnbaum, M. H. Problems with so-called "direct" scaling methods. In J. T. Kuznicki, R. A. Johnson, & A. F. Rutkiewic (Eds.), *Sensory methods: Problems and approaches to hedonics* (ASTM STP773). Philadelphia, PA: American Society for Testing and Materials, 1982. (a)

Birnbaum, M. H. Controversies in psychological measurement. In B. Wegner (Ed.), *Social attitudes and psychophysical measurement*. Hillsdale, N.J: Erlbaum, 1982. (b)

Birnbaum, M. H., Parducci, A., & Gifford, R. K. Contextual effects in information integration. *Journal of Experimental Psychology*, 1971, **88**, 158-170.

Birnbaum, M. H., & Viet, C. T. Scale convergence as a criterion for rescaling: Information integration for difference, ratio, and averaging tasks. *Perception & Psychophysics*, 1974, **15**, 7-15.

Helson, H. *Adaptation level theory*. New York: Harper & Row, 1964.

Johnson, D. M., & Mullally, C. R. Correlation-and-regression model for category judgments. *Psychological Review*, 1969, **76**, 205-215.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. *Foundation of measurement*. New York: Academic Press, 1971.

Mellers, B. A. Equity judgment: A revision of Aristotelian views. *Journal of Experimental Psychology: General*, 1982, **111**, 242-270.

Mellers, B. A. Evidence against "absolute" scaling. *Perception & Psychophysics*, 1983, **33**, 523-526.

Mellers, B. A., & Birnbaum, M. H. Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception & Performance*, 1982, **8**, 582-601.

Mellers, B. A., & Birnbaum, M. H. Contextual effects in social judgment. *Journal of Experimental Social Psychology*, 1983, **19**, 157-171.

Parducci, A. Range-frequency compromise in judgment. *Psychological Monographs*, 1963, **77**(2, Whole No. 565).

Parducci, A. Contextual effects: A range-frequency analysis. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 2). New York: Academic Press, 1974.

Parducci, A. Category ratings: Still more contextual effects! In B. Wegner (Ed.), *Social attitudes and psychophysical measurement*. Hillsdale, N.J: Erlbaum, 1982.

Parducci, A., & Perrett, L. Category rating scales: Effects of relative spacing and frequency of stimulus values. *Journal of Experimental Psychology*, 1971, **89**, 427-452.

Poulton, E. C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, 1968, **69**, 1-19.

Restle, F., & Greeno, J. G. *Introduction to mathematical psychology*. Reading, Mass: Addison-Wesley, 1970.

Stevens, S. S. Measurement and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

Zwislocki, J. J. Absolute and other scales: The question of validity. *Perception & Psychophysics*, 1983, **33**, 593-594.

Zwislocki, J. J., & Goodman, D. A. Absolute scaling of sensory magnitudes: A validation. *Perception & Psychophysics*, 1980, **28**, 28-38.

NOTE

1. Although Zwislocki does not provide a theory of the "biases" of Mellers (1983), he does speculate about two causes of "bias." First, he states that the stimulus range of Mellers was too small. But, if the stimuli of Mellers (1983) did not vary in sub-

jective darkness, why did Mellers obtain large main effects of the stimuli, let alone effects of the context across conditions? A small stimulus range would have biased the experiment toward the null hypothesis that judgments were invariant.

Second, Zwislocki wonders whether contextual effects found by Mellers may be limited to experiments using simultaneous presentation of the stimuli. Parducci (1963, 1974, 1982) and Parducci and Perrett (1971) found contextual effects due to the stimulus distribution with sequential presentation. Furthermore, they have found contextual effects with simultaneous and sequential presentation of the stimuli.

(Manuscript received July 29, 1983;
accepted for publication July 29, 1983.)