DISCUSSION

M. H. Birnbaum[1] *(written discussion)*—Moskowitz contends that magnitude estimation procedures are more sensitive for detecting stimulus differences than the procedure of category rating. He had groups of subjects judge stimuli by either procedure, yielding a data array of subjects by stimuli for each procedure. Each matrix was decomposed by analysis of variance. Moskowitz found that the F-ratio (mean squares for stimuli divided by mean squares for the stimuli by subjects' interaction) is often greater for magnitude estimation than for his category rating procedure.

This result runs contrary to other findings, however. For example, the same type of F-ratio was computed for 12 magnitude estimation conditions and 6 category rating conditions of Mellers and Birnbaum. The subjects in their studies judged the darkness of dot patterns using different procedures. The number of subjects was equalized by taking the first 22 subjects in each condition. It was found, contrary to Moskowitz, that category ratings produce larger F-ratios than magnitude estimations. Of the 12 Fs for magnitude estimation, 6 were less than 80, and only 2 were greater than 120. However, all of the 6 Fs for category ratings were greater than 80, and 5 of the 6 were greater than 120. Thus, by the criterion of Moskowitz, category ratings are preferable to magnitude estimations.

It seems plausible that the findings of Moskowitz are related to the well-known "magic number seven plus or minus two" phenomenon; information transmission is optimal when the number of response categories is about seven. When asked to rate three "good" cakes, subjects might very well rate them all "good," but they might assign different numbers when using magnitude estimation. With the procedure of Moskowitz, therefore, subjects may use only two or three categories. It seems likely that the conclusions of Moskowitz would be reversed if the following category rating procedure were adopted: (*a*) Ask the subjects to experience all of the stimuli (taste all three cakes, or whatever) and (*b*) Instruct them to assign "1" to the worst (lowest) stimulus and "9" to the best (greatest) stimulus and to rate the others in between according to subjective value.

This procedure ensures that more than two or three categories will be used for the stimuli, and thus, the "magic number," seven plus two (9), categories will be used. Given the results of Mellers and Birnbaum, it seems likely that this category rating procedure will yield higher F-values than either of the two "direct" scaling procedures compared by Moskowitz.

When the number of stimuli are small, stimuli can be judged in pairs in-

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

stead of singly without a large number of trials. If the aim is to detect small differences between stimuli, it seems reasonable to present stimuli in pairs and ask subjects to judge the "differences" between the stimuli. For example, subjects would taste cakes A, B, and C, and they would be asked to rate the "differences" between A and B, B and C, and A and C, indicating strength of preference. They could be asked to identify the worst and best cake, to "call that difference 100," and to "judge other differences relative to that difference."

Such an experiment provides not only the rank order of the stimuli, but also the rank order of the differences among the stimuli, permitting construction of a quantitative scale of subjective value. Further advantages of the subtractive model are given in Birnbaum's chapter in this publication.[2]

H. R. Moskowitz (author's closure)—Dr. Birnbaum's suggestions, 1 and 2, imply that we allow the subjects to experience the entire range of products, and then force the poorest product to achieve a value, 1, and the best to achieve a value, 9. This procedure totally falls apart, if panelists are allowed to test a random subset of all products, as is often done. Panelists should not have to assess all products first, before using the scale. Rather, if the scale is valid, we should be able to acquire good, sensitive data, if the panelists make judgments in the sequence that he or she tests the products. One wonders whether this desired "preview" of all of the stimuli somehow alerts the panelists in Dr. Birnbaum's study to be ready to modify their innate use of the scale in order to accommodate the new set of stimuli. How comparable, in turn, are ratings from quite different sets of stimuli on the same scale, if the preview is necessary to adjust the scaling behavior of subjects, for each new stimulus set?

[2]Birnbaum, M. H., this publication, pp. 34–48.