Evidence against "absolute" scaling

BARBARA A. MELLERS

University of California, Berkeley, California

Zwislocki and Goodman (1980) argue that there exists an "absolute coupling" of numbers to sensation magnitudes and conclude that, when subjects are left unconstrained by a designated stimulus-number pair, they may use this "absolute" scale. The purpose of the studies reported here was to test whether Zwislocki and Goodman's (1980) absolute scaling procedure reduces contextual effects due to variations in the stimulus spacing. It was found that magnitude estimations vary as a function of the stimulus spacing, regardless of whether subjects are instructed to use a standard and modulus, and, furthermore, that category ratings yield effects of the stimulus spacing comparable to those obtained with magnitude estimations. It is argued that removing the so-called constraints of a standard and modulus does not yield an "absolute" scale of sensation. The absolute scaling procedure increases response variability and thereby lowers the power of a test for contextual effects.

Zwislocki and Goodman (1980) argue that, if subjects are allowed to choose their own standards and moduli in magnitude estimations, they may use an "absolute" scale that is invariant with respect to contextual effects due to such factors as the first stimulus presented, the range of stimuli, the range of numbers, and previous experience. Their procedure is reminiscent of earlier views in scaling that argued against "constraints" on the subject. For example, Stevens and Galanter (1957) argued that magnitude estimations should be preferred to category ratings, because magnitude estimations, unlike category ratings, do not restrict the subject's range of responses. When it became apparent that magnitude estimations were susceptible to contextual effects due to variations in the standard, the modulus, the stimulus range, and the stimulus spacing (Poulton, 1968), procedures were suggested to further increase subjects' "freedom," such as allowing them to select their own standards (Poulton, 1979). Zwislocki and Goodman (1980) suggest that subjects should select both the standard and the modulus. However, they go one step further by claiming that with such a procedure, an "absolute" scale may emerge.

Evaluating the Magnitude of Contextual Effects

The results of Zwislocki and Goodman (Figures 4-12, 1980) show evidence of contextual effects. However, Zwislocki and Goodman are impressed with the stability of the results and conclude that the data support their "absolute" scale hypothesis. To

evaluate the magnitude of contextual effects, it is necessary to provide a frame of reference for comparison. In other words, contextual effects should be eliminated or at least reduced using the Zwislocki and Goodman (1980) procedure relative to other procedures. Zwislocki and Goodman (1980) do not vary the instructions and thus provide no frame of reference for evaluating the magnitude of their effects.

The present studies were designed to investigate the magnitude of contextual effects due to variations in the stimulus spacing for the absolute scaling procedure relative to other procedures. A set of common stimuli previously used by Mellers and Birnbaum (1982) was presented in the context of two stimulus distributions. In Study 1, the magnitude of contextual effects for the two distributions was investigated using two magnitude estimation procedures. In one condition, a standard and modulus were designated, and in the other condition, the Zwislocki and Goodman instructions were used. If Zwislocki and Goodman's procedure produces an "absolute" scale, then, with the absolute scaling procedure, contextual effects should be reduced relative to the other condition.

In Study 2, the magnitudes of contextual effects were compared using two different tasks, category rating and magnitude estimation. Study 2 was designed to provide a larger frame of reference for evaluating the magnitude of contextual effects by manipulating not only the instructions for a task, but also the tasks.

METHOD

The author thanks Michael H. Birnbaum for many valuable comments and suggestions and Elizabeth Prather for assistance with the data collection. Request for reprints should be sent to Barbara Mellers, Department of Psychology, University of California, Berkeley, California 94720.

Subjects made magnitude estimations of the subjective darkness of dot patterns. There were two different stimulus spacing conditions in each study. In Study 1, the standard and the modulus were not specified and the association of numbers to stimulus magnitudes was left entirely to the subjects, following the pro-

cedure of Zwislocki and Goodman (1980). In Study 2, judges were instructed to use a specified standard and modulus. Each judge served in only one of the four conditions.

Instructions

Study 1. Instructions were identical to those of Zwislocki and Goodman (1980, p. 38), except for the stimulus dimension. The instructions read as follows: "In this experiment, we would like to find out how dark various dot patterns appear to you. Your task will be to assign a number to every dot pattern in such a way that your impression of how large the number is matches your impression of how dark the dot pattern is. You may use any positive numbers that appear appropriate to you—whole numbers, decimals, or fractions. Don't think of physical units of measurement, such as the actual number of dots, and do not worry about running out of numbers—there will always be a smaller number than the smallest you use and a larger one than the largest you use."

Study 2. In the second study, the darkest pattern was designated as the standard and assigned the modulus "100." Examples given were as follows: "1" = 1% as dark as the standard dot pattern, "20" = 20% as dark as the standard, "40" = 40% as dark as the standard, "60" = 60% as dark as the standard, "80" = 80% as dark as the standard, "100" = equal in darkness to the standard. Judges were encouraged to use the numbers in between the examples.

Stimuli

The 11 stimuli in each condition consisted of 25-mm squares containing irregular dot patterns of 1-mm solid dots. The number of dots was varied according to either a positively skewed or negatively skewed distribution. Six stimuli were common to both distributions and contained 12, 18, 27, 40, 60, and 90 dots. In the positively skewed distribution, five additional context stimuli had 14, 15, 16, 21, and 23 dots. In the negatively skewed distribution, five context stimuli had 47, 51, 70, 74, and 77 dots. The stimuli and presentation format are shown in Figure 2 of Mellers and Birnbaum (1982).

Participants

Judges were University of California, Berkeley, undergraduates who were enrolled in psychology courses. There were 100 and 98 judges in the two conditions of Study 1; 27 and 25 judges served in Study 2. In Study 1, 14 judges who failed to follow instructions (by using nonpositive numbers) were not included in the analyses.

RESULTS AND DISCUSSION

Study 1: Contextual Effects with "Absolute Scaling" Instructions

Panel A of Figure 1 shows geometric means for the six common dot patterns. Judgments are plotted against the number of dots, which are spaced according to the log of the actual number of dots, Φ . The upper curve connects mean judgments from the positively skewed distribution; the lower curve connects means from the negatively skewed distribution. Brackets represent plus and minus 1/12.5 of a standard deviation in all three panels.

If Zwislocki and Goodman's procedure allows an "absolute" scale to emerge, then the same stimuli should receive the same average judgments, irrespective of the other stimuli with which they are presented. However, Figure 1 shows that the same stimuli receive different responses in different contexts; therefore, it appears that subjects do not use an "absolute" scale even when they select their own stan-

Figure 1. Magnitude estimations of six common dot patterns. Panel A shows geometric means of subjects who were given the "absolute" scaling instructions of Zwislocki and Goodman. Panels B and C show arithmetic means for subjects who were told to use the 12-dot pattern as the standard and "100" as the modulus. The two curves in Panel B show results when instructions included an example response as high as "900"; Panel C shows results when the highest example was "300." Brackets represent plus and minus 1/12.5 of standard deviation in all three panels. (POS = positively skewed condition; NEG = negatively skewed condition.)

dards and moduli. Furthermore, there is no evidence for a "ratio" scale. Not only single responses, but also ratios of responses vary with contextual manipulations. Notice that the response to the 90-dot pattern divided by the response to the 12-dot pattern is only 5.25 ("21"/"4") in the positively skewed condition, whereas, in the negatively skewed condition, this ratio is about 7 ("14"/"2").

To check whether contextual effects occurred for judges who used both large and small ranges, the subjects were subdivided according to their largest response. For all subgroups, it was found that judgments from the positively skewed distribution were negatively accelerated relative to those from the negatively skewed distribution.

Panels B and C of Figure 1 show geometric mean magnitude estimations of the same stimuli with the 12-dot pattern as the standard and "100" as the modulus (from Mellers & Birnbaum, 1982). The two curves in Panel B were based on instructions that included examples that went up to 900 (e.g., "If the dot pattern seems 9 times as dark as the standard, respond 900"). The two curves in Panel C were obtained using instructions with examples that went up to 300 (i.e., "If the dot pattern seems 3 times as dark, respond 300"). Panels B and C show that the examples provided in the instructions, as well as the stimulus spacing, affect magnitude estimations. When the examples range as high as "900," the subjects use larger numbers on the average than when the largest example is only "300." Robinson (1976) also found that examples affect magnitude estimations. Taken together, Panels A, B, and C show that the effect of the stimulus spacing is similar regardless of whether examples are provided in the instructions.

Study 2: Contextual Effects with Bounded Scales

Panel A of Figure 2 shows arithmetic mean magnitude estimations in the conditions in which subjects were asked to use the 90-dot pattern as the standard and "100" as the modulus. The effects of the stimulus spacing appear similar to those in Study 1, when the standard and the modulus were not presented. Note that the curve from the positively skewed condition lies above the curve for the negatively skewed condition and is negatively accelerated relative to the negative condition, as in Figure 1. Furthermore, these magnitude estimation results are similar to the results obtained using a category rating with a 100-point scale, shown in Panel B of Figure 2 (from Mellers & Birnbaum, 1982).

Mellers and Birnbaum (1982) found that the effects of the stimulus spacing were similar for magnitude estimations and category ratings. The main difference between the two procedures was that with magnitude estimations, curves representing judgments of the common stimuli from negatively and positively skewed contexts did not rejoin at the upper ends, whereas with category ratings, curves from the two contexts typically did rejoin.

Magnitude estimations, unlike category ratings, can be either bounded or unbounded at the upper end, depending on the standard. The present results show that when magnitude estimations and category ratings are bounded at both ends, the curves rejoin. Montgomery (1975) also found that when magnitude estimations were bounded, they resembled category ratings.

Increasing Response Variability

Moskowitz (1982) argued that magnitude estimations are more sensitive to stimulus differences than

are category ratings. He based this claim on a comparison of F ratios for the main effect of stimuli relative to the stimuli $\times$ subjects interaction.¹ Similar F ratios were computed for the six common dot patterns in the 10 conditions of Figures 1 and 2. All of the F ratios were based on 22 subjects for comparability. However, four Fs were computed for each of the two stimulus spacing conditions with the absolute scaling procedure. The largest $F(5,105)$ was 13.04 for the conditions using Zwislowski and Goodman's instructions, while the smallest $F(5,105)$ was 61.28 for all of the remaining conditions. In summary, it appears that the Zwislowski and Goodman procedure is the least sensitive to stimulus differences, according to the criterion of Moskowitz (1982).

Since the Zwislowski and Goodman (1980) procedure enhances individual differences, it reduces the power of the test for contextual effects. It may be easier to conclude that the magnitude of contextual effects is nonsignificant with the Zwislowski and Goodman instructions, even if it is of the same magnitude as with other procedures. Despite the increased subjects variability with the Zwislowski and Goodman procedure, in an analysis of variance on the log judgments of the common stimuli, the main effect of context and the context $\times$ stimuli interaction were statistically significant [$F(1,194) = 13.62$ and $F(5,194) = 10.62$, respectively]. Thus, contextual effects were significant, even though the power of the test was reduced with the absolute scaling procedure.

CONCLUSIONS

Zwislowski and Goodman (1980) argue that when subjects are left unconstrained, they tend to use an "absolute" scale. By definition, an "absolute" scale implies a scale that is invariant with respect to the context. These experiments use the procedure of Zwislowski and Goodman and show contextual effects due to variations in the stimulus spacing. Furthermore, these contextual effects resemble those for magnitude estimations with a designated stimulus-number pair and those for category ratings.

An alternative view, taken by Birnbaum (1974, 1975, 1982a, 1982b) and Mellers and Birnbaum (1982, 1983), is that all judgments are relative and occur in a context. When judging a stimulus, the subject compares the relative position of the stimulus within a context. Birnbaum (1975, 1982a, 1982b) points out that the subject is actually influenced by two contexts: (1) an outside-the-laboratory context based on the subject's past and present experiences, and (2) an inside-the-laboratory context determined by an experimenter's choice of stimuli, procedure, etc. Both contexts should have an effect on judgments, according to this relativity theory (Mellers,

Figure 2. Panel A shows magnitude estimations for the six common stimuli in the positively and negatively skewed conditions. The 90-dot pattern was the standard and was assigned a value of "100." Panel B shows category ratings on a 100-point scale for the same six stimuli. Brackets represent plus and minus one half of a standard error.

1982; Mellers & Birnbaum, 1982). When the experimenter relaxes the "constraints" of an assigned modulus and standard, contextual effects outside the laboratory may determine the subject's responses to a greater extent than perhaps they might in a more restricted task. But the experimental context, that is, the stimulus distribution, response procedures, etc., also influences the subject's judgments.

The results shown in Figures 1 and 2 are in the direction predicted by a particular contextual theory, namely, range-frequency theory (Parducci, 1968, 1974, 1982). Range-frequency theory asserts that the response to a stimulus is a compromise between the rank order of the stimulus in the distribution (frequency component) and the position of the stimulus relative to the endpoints (range component). According to this theory, judgments of the common stimuli in the positively skewed distributions of Figures 1 and 2 should be negatively accelerated relative to those from the negatively skewed distributions. Both the magnitude estimations and the category ratings are consistent with this prediction.

The relativity viewpoint can also account for the results of Zwislocki and Goodman (1980). When subjects use a less "constrained" procedure, such as that of Zwislocki and Goodman, then responses may become more influenced by the context outside the laboratory. It seems entirely possible that, in everyday experiences, there is some consistency about what would be called a "loud" noise or a "big" number. Thus, the more "invariant" results of Zwislocki and Goodman could be due to some consistency in subjects' outside-the-laboratory context rather than their use of an "absolute" scale. Relativity theory (Mellers, 1982; Mellers & Birnbaum, 1982) asserts that this context, as well as the laboratory context, should influence judgments.

In conclusion, it appears that contextual effects cannot be "avoided" by selecting a procedure that gives the subject "freedom." Different procedures affect the context differently and thereby affect judgments differently, but no procedure should be regarded as context free. The supposed freedom of the subject in magnitude estimations may actually be guided by the stimulus range, the stimulus spacing, and other procedures determined by the experimenter. When no standards or moduli are designated, subjects may be relatively more influenced by outside-the-laboratory contexts. They may match "big" numbers or "loud" tones according to the relative positions of the numbers and tones in their respective distributions, both inside and outside the laboratory. Thus, the procedure of Zwislocki and Goodman (1980) does not appear to produce an "absolute" scale. However, it does increase responsible variability and reduce the power of the test for contextual effects.

REFERENCES

- BIRNBAUM, M. H. Using contextual effects to derive psychophysical scales. *Perception & Psychophysics*, 1974, 15, 89-96.
- BIRNBAUM, M. H. Expectancy and judgment. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1975.
- BIRNBAUM, M. H. Controversies in psychological measurement. In B. Wegner (Ed.), *Social attitudes and psychophysical measurement*. Hillsdale, N.J.: Erlbaum, 1982. (a)
- BIRNBAUM, N. H. Problems with so-called "direct" scaling. In J. T. Kuznicki, R. A. Johnson, & A. F. Rutkiewicz (Eds.), *Selected sensory methods: Problems and approaches to hedonics* (ASTM STP773). Philadelphia: American Society for Testing and Materials, 1982. (b)
- MELLERS, B. A. Equity judgment: A revision of Aristotelian views. *Journal of Experimental Psychology: General*, 1982, 111, 242-270.
- MELLERS, B. A., & BIRNBAUM, M. H. Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception and 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.
- MONTGOMERY, H. Direct estimation: Effect of methodological factors on scale type. *Scandinavian Journal of Psychology*, 1975, 16, 19-29.
- MOSKOWITZ, H. R. Utilitarian benefits of magnitude estimation scaling for testing product acceptability. In J. T. Kuznicki, R. A. Johnson, & A. F. Rutkiewicz (Eds.), *Selected sensory methods: Problems and approaches to hedonics* (ASTM STP773). Philadelphia: American Society for Testing and Materials, 1982.
- PARDUCCI, A. The relativity of absolute judgment. *Scientific American*, 1968, 219, 84-90.
- 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.
- POULTON, E. C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, 1968, 69, 1-19.
- POULTON, E. C. Models for biases in judging sensory magnitude. *Psychological Bulletin*, 1979, 86, 777-803.
- ROBINSON, G. H. Biasing power law exponents by magnitude estimation instructions. *Perception & Psychophysics*, 1976, 19, 80-84.
- STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, 54, 377-411.
- ZWISLOCKI, J. J., & GOODMAN, D. A. Absolute scaling of sensory magnitudes: A validation. *Perception & Psychophysics*, 1980, 28, 28-38.

NOTE

1. Mellers and Birnbaum (1982) computed such F ratios for each of 6 conditions of category ratings and 12 conditions of magnitude estimations. The F ratios for category ratings tended to be larger than those for magnitude estimations. All 6 category rating conditions had $F(5,105)$ greater than 80, and 5 of the 6 Fs were greater than 120. But for the 12 magnitude estimation conditions, only 6 of the 12 had $F(5,105)$ greater than 80, and 2 Fs were greater than 120. Thus, the relative variability due to the subjects $\times$ stimuli interaction was greater for magnitude estimations than for category ratings, contrary to the claim of Moskowitz (1982).

(Manuscript received July 6, 1982;
revision accepted for publication January 26, 1983.)