

Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments

LESTER E. KRUEGER
The Ohio State University, Columbus, Ohio

In previous studies, modalities with a higher Weber fraction have tended to have a lower power-function exponent. Within a modality, however, the Weber fraction and power-function exponent for individual subjects were unrelated, and the present study largely confirms this finding for the numerosity dimension. More important than discriminability in the judgment of numerosity were cognitive factors. A single feedback trial considerably reduced intersubject variability on the magnitude-estimation exponent, although it failed to eliminate individual differences completely (precue and postcue exponents correlated significantly, $r = +.50$). Intra-subject variability, by contrast, seemingly did not involve the underlying exponent. As in previous studies, numerosity generally was underestimated and the power-function exponent was 1.08 for magnitude production and .80 for precue magnitude estimation. Contrary to previous results, however, males and females did not differ in exponent, perhaps because the present procedure allowed self-selection of individuals more interested in numerosity tasks.

People tend to underestimate the number of elements present in a display, and numerosity judgments typically form a power function (Stevens, 1957) with an exponent of about .85 (Indow & Ida, 1977; Krueger, 1972, 1982). We do not know whether the underestimation is due to cognitive factors or to perceptual or sensory ones. If it is due to cognitive factors, the underestimation might be dispelled by altering the instructions on magnitude estimation (M. Teghtsoonian, 1965) or by providing knowledge of results (Minturn & Reese, 1951). On the other hand, if the underestimation reflects an inability to discriminate differences in the number of elements present in a display, then neither instructions nor feedback should affect it.

Some results indicate that the exponent for a particular modality generally depends more on perceptual or sensory factors than on cognitive ones, being related most fundamentally to the discriminability of stimuli in that modality (Auerbach, 1971; Heinemann, 1978). Baird (1970; Baird, Romer, & Stein, 1970; Graf, Baird, & Glesman, 1974) and R. Teghtsoonian (1971) found a negative correlation between the size of the exponent and the Weber fraction (or related measure) for a particular modality or condition. This finding is consistent with Fechner's law (Auerbach, 1971; R. Teghtsoonian, 1971), because if subjective

magnitude depends on the number of just noticeable differences (jnDs) above the absolute threshold, as Fechner claimed, then the larger the Weber fraction, the more the physical magnitude must be changed to produce a particular change in the number of jnds and thus the subjective magnitude, just as Baird and R. Teghtsoonian found. One may ask whether this relationship between the Weber fraction and the exponent, which holds across modalities, also will hold for individual subjects within a particular modality. In the present study, the exponents for numerosity, obtained with magnitude production and magnitude estimation for individual subjects, were compared with the percentage of errors the same subjects made in judging which of two simultaneously presented sheets contained more Xs. Graf et al. (1974) also tested the association between the exponent and the jnd or Weber fraction to determine if it would hold on individual differences. Although negative, as predicted by the theory linking the exponent with discriminability, their correlations, which ranged from $-.01$ to $-.25$, were insignificant. The present study used a larger sample (100 rather than 48 subjects) and a different modality (numerosity rather than light intensity) in order to provide another test of the theory.

Perceived numerosity is a higher level, less sensory dimension than loudness or brightness, so it ought to be more susceptible to cognitive influence. ("Cognitive" refers here to the interpretation stage that follows the perceptual and sensory process, and that assigns a number or some other response to the input.) Minturn and Reese (1951), who presented 1 to 210 dots, found that feedback on numerosity judgments reduced variability, producing a convergence

The author is grateful to Constance Stewart for assistance in collecting and tabulating the data, and to Martha Teghtsoonian for helpful comments on an earlier version of this report. Requests for reprints should be sent to Lester E. Krueger, Human Performance Center, Ohio State University, 404-B West 17th Avenue, Columbus, OH 43210.

in the judgments of under- and overestimating subjects. Some reduction in variability was still evident 8 months later. The effect of feedback was general, because a shift occurred even on stimulus levels for which feedback had not yet been given. If feedback affected the underlying psychophysical function rather than just the number assigned to a particular stimulus level, then much of the effect may have occurred on the very first feedback trial. To test this further in the present study, a single feedback trial was given between the two halves of the magnitude estimation task. If perceived numerosity depends on cognitive factors, this feedback ought to reduce variability among subjects. It was also expected to increase the magnitude-estimation exponent for most subjects, moving it closer towards a veridical value of 1.0.

In order to compare variability within subjects with that between subjects, each subject received two complete, successive replications of all conditions in magnitude production and magnitude estimation. Krueger (1982) obtained single magnitude estimations of numerosity from each of 100 subjects at each of eight stimulus levels (number of Xs presented). He found that the intersubject standard deviation (SD) of logarithmically transformed responses at a particular stimulus level increased as stimulus level increased. Since each subject's scale was anchored at a numerosity value just above 6 (the largest number readily subitized, according to Kaufman, Lord, Reese, & Volkman, 1949), the increase in intersubject SD apparently reflected variation in the size of the individual power-function exponents rather than variability in the scale factor or noise in retrieving particular responses (Weber's law). The present study examined whether much of the variation in the power-function exponent occurs within individual subjects.

Previous experiments in my laboratory never presented fewer than 25 dots or Xs (Krueger, 1972, 1982). Overestimation, rather than underestimation, is sometimes found when fewer than 25 elements are presented for magnitude estimation, perhaps reflecting a regression effect or contraction bias (Poulton, 1979; Stevens, 1966). Minturn and Reese (1951) found under- or overestimation, depending on whether more or fewer than 20 (but more than 5) elements were presented. Kaufman et al. (1949) also found overestimation in the lower region, but the crossover point in their case was 11, not 20. The present study included displays of 10, 15, and 20 Xs in the magnitude estimation task in order to test further whether overestimation occurs in the lower stimulus region.

The tendency of females to underestimate numerosity to a greater extent than males (Krueger, 1982) was examined again, using the present battery of tasks and conditions to determine the basis of the sex difference.

METHOD

Subjects

One hundred Ohio State University undergraduates (50 male, 50 female) participated in a 1-h experimental session in order to receive credit in an introductory psychology course. All subjects had normal vision; in case of doubt, a Snellen test was administered, and 20/40 performance or better was required.

Each male subject was matched with a particular female subject, so all counterbalancing, reversing, and randomization of trials mentioned below applied equally to the two sets of subjects. The subjects volunteered specifically for a set of numerosity tasks. In the first quarter of data collection, 39 males, but only 17 females, were recruited. In subsequent quarters, recruiting had to be restricted at times so that the final sample would include 50 females.

Test Stimuli and Procedure

Each computer-generated white printout sheet (37.78 cm wide \times 27.94 cm high) contained a 35-column \times 25-row matrix (26.04 cm wide \times 20.48 cm high) of possible X positions. Each successive set of 5 rows contained exactly (magnitude production, magnitude estimation) or approximately (discrimination) 20% of the Xs specified, which were distributed randomly over the 175 positions in the 5 rows (Krueger, 1972, 1982). All sheets, whether in a fan-folded stack (magnitude production, discrimination) or separated (magnitude estimation), were shown in vertical orientation. Unlike those used by Krueger (1982), the present printout sheets were unlined.

An invariant order of tasks (magnitude production, magnitude estimation, and discrimination) was used, so that the cue stimulus presented in the magnitude estimation task could not bias magnitude production. In magnitude production and magnitude estimation, the basic set of trials was presented twice, so that variability within as well as between subjects could be measured.

Magnitude production. A single set of 173 sheets was prepared, with the Xs increased (in steps of 5) from 15 to 875, which fully filled the 35 \times 25 matrix. A blank sheet preceded each printed sheet, so that the subjects could turn the pages of the fan-folded stack from right to left, just as in a regular book, and observe each printed display in the same (right-side) location.

In each set of four trials, the subjects looked for sheets containing 25, 50, 100, and 200 Xs. The subjects started turning the pages from the part of the stack with either a low number of Xs (low start) or a higher number of Xs (high start). The two low-start replications preceded the two high-start ones for half of the subjects; the reverse order was used for the other half. Different orders of the four assigned numbers were used for the four replications, and the 24 possible orders were used approximately equally often across subjects.

The subjects were asked to bracket their judgments, that is, to leaf beyond the first sheet that appeared to match the target number of Xs, and then come back so as to home in on the best possible match. The subjects were cautioned not to try to count the Xs or to compute mentally how many there must be on a given page, but to base their judgments solely on how many Xs appeared to be present.

Magnitude estimation. Each set of 12 stimuli contained the following number of Xs: 10, 15, 20, 25, 35, 50, 75, 100, 150, 200, 300, and 400. Six different sets of stimuli were printed, and 12 different random orderings of stimuli within sets were devised. Four different sets and orderings were used for the four replications (two precue, two postcue) presented to a particular subject.

The subjects were asked to take only a second or two to form an impression of how many Xs were present, and they again were cautioned not to try to count or to compute mentally the number of Xs present. When the cue stimulus was shown, the subject first was asked how many Xs he or she thought it contained and then was told that it contained 200 Xs. The subject was asked to bear

this feedback in mind when making further magnitude estimations.

Discrimination. A single set of computer printout sheets was prepared; these contained two printed sheets, followed by two blank sheets, for each of 96 trials. As the pages of the fan-folded stack of sheets were turned from right to left, the two printed sheets for a particular trial were exposed. One sheet (standard stimulus) contained the standard number of Xs, whereas the other (comparison stimulus) differed from the standard by $-4Y$, $-3Y$, $-2Y$, $-Y$, Y , $2Y$, $3Y$, and $4Y$, where Y took the values 1, 3, and 9, respectively, for standards of size 25, 100, and 400. Thus, there were eight comparisons for each of the three standards, making 24 basic stimuli. Each of the four successive blocks of 24 trials (total of 96 trials) replicated the 24 basic conditions. In each block, the more numerous sheet appeared on the left for 12 trials, and on the right for the other 12 trials. Across blocks, the more numerous sheet for each of the 24 basic stimuli appeared twice on the left and twice on the right.

The subjects had to decide which sheet (left or right) contained more Xs. They were told that one sheet always contained more Xs, and that they should choose which one that was even if they felt they were only guessing. They were cautioned again not to try to count the Xs and to spend only a few seconds making each decision. Half of the subjects went through the 96 trials from one end of the stack; the other half did so from the other end.

RESULTS AND DISCUSSION

In all tasks, functions were fitted to data using the least squares criterion. In magnitude production and magnitude estimation, data were logarithmically transformed before means and SDs were computed. Thus, both within and between subjects, the means were geometric. The group means were plotted on log-log coordinates (magnitude production in Figure 1 and magnitude estimation in Figure 3), and the power function will be referred to below in its logarithmic form, $\log R = \log a + n \log S$, in which R is the number of Xs reported or assigned, S is the number of Xs presented or produced, $\log a$ is the scale factor (intercept on log-log coordinates), and n is the exponent (slope on log-log coordinates).

Variability of the response (reported or produced) for a particular stimulus size (presented or assigned) was computed both for the group and for individual subjects. For the group SD, each subject provided one datum (i.e., the mean log value for the two replications). The group SD reflects both inter- and intra-subject variability, with the latter reduced by a factor of $\sqrt{2}$, due to the averaging across two trials. With the scale anchored at a value below 10 (Krueger, 1982), the group SDs (solid horizontal or vertical lines in Figures 1 and 3 below) ought, as stimulus size increased, to remain constant on a log-log plot, if they reflected variability in the retrieval of a particular response (Weber's law) or in the scale factor; they ought to increase if they reflected variability in the exponent. The latter result by itself, which is what Krueger (1982) found, would not indicate whether the exponent varied within or between subjects. Therefore, the (mean) intrasubject SDs (dotted horizontal or vertical lines in Figures 1 and 3 below) were computed, based on the two replications at each stimulus

size for each subject. For ease of comparison with the group SDs, they are shown in Figures 1 and 3 as deviations from the overall mean log response.

Magnitude Production

Due entirely to the negative scale factor ($-.31$ for low start, $-.27$ for high start)—since the exponent actually exceeded 1.0 for both low start (1.12) and high start (1.03)—the subjects underestimated the number of Xs present, as shown in Figure 1. These exponents exceeded those for magnitude estimation, but the stimulus range was smaller than that for magnitude estimation; the difference thus might be a regression effect (Stevens, 1966) or a range effect (R. Teghtsoonian & M. Teghtsoonian, 1978).

Exponents also were computed for each subject, as shown in Figure 2, both for low start ($SD = .293$) and high start ($SD = .310$). The high-start scores deviated significantly from the normal distribution (shown as a dotted line in Figure 2) ($D = .136$, $p < .05$; Kolmogorov-Smirnov one-sample test; Siegel, 1956), but the low-start scores did not ($D = .116$, $p < .15$). The subjects differed fairly consistently between replications on low start [scale factor, $r(98) = +.29$, $p < .01$; exponent, $r(98) = +.56$, $p < .001$], between

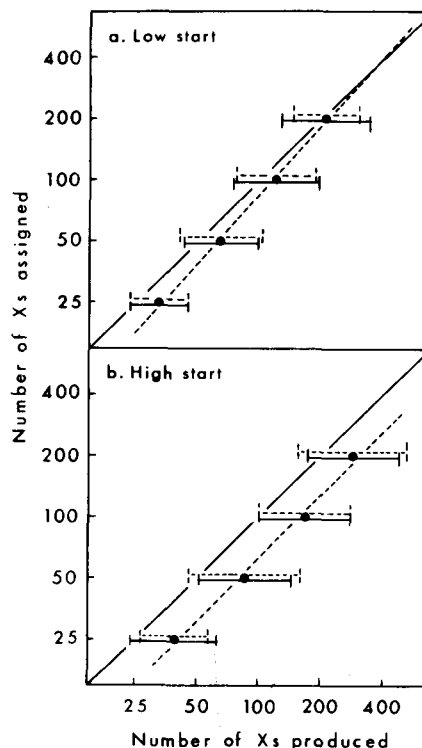


Figure 1. Magnitude production: Geometric mean number of Xs produced (filled circles) by number of Xs assigned, on log-log coordinates, for low-start and high-start conditions. The solid diagonal line (veridical function) has a slope of 1.0. The solid horizontal line represents the group standard deviation, and the dotted horizontal line represents the mean intrasubject standard deviation across the two replications.

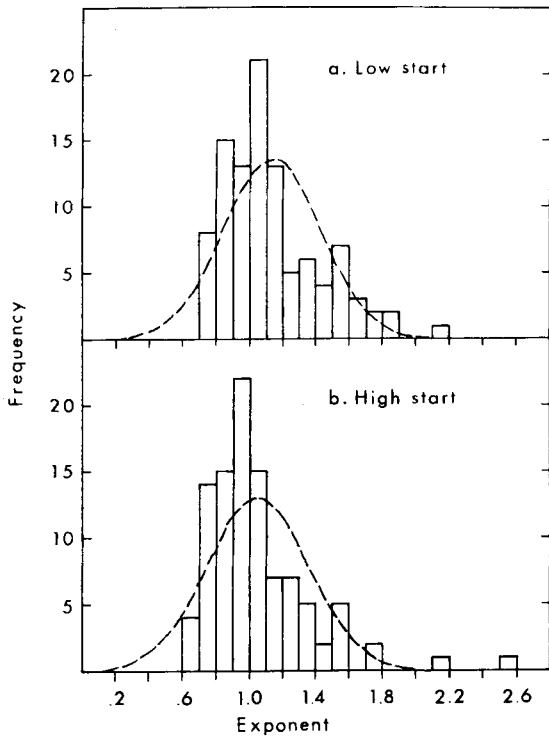


Figure 2. Magnitude production: Distribution of exponents for individual subjects. The dotted line depicts a normal distribution.

replications on high start [scale factor, $r(98) = +.12$, n.s.; exponent, $r(98) = +.52$, $p < .001$], and between low start and high start [scale factor, $r(98) = +.38$, $p < .001$; exponent, $r(98) = +.50$, $p < .001$].

The group SDs (solid horizontal lines in Figure 1) steadily increased on low start (slope of .09), but not on high start (slope of .01), as the log number assigned increased, whereas the intrasubject SDs (dotted horizontal lines in Figure 1) did not become progressively larger, but varied nonmonotonically with number of Xs assigned. Thus, there was inter- but not intrasubject variability on the exponent.

Magnitude Estimation

As shown in Figure 3, subjects generally underestimated the number of Xs present, especially in the precue condition. The exponent for the postcue condition was only slightly higher than that for the precue condition (.82 vs. .80), however, and they would have been identical (.76) for both conditions if the lowest three data points (10, 15, 20), which fell exactly along the veridical function in the precue condition, had been excluded. More important was the increase in the scale factor, from .25 (precue) to .29 (postcue). As a result, the crossover to underestimation occurred at a higher level in the postcue condition. Since the crossover point was stable within each condition (15.28 and 19.48 Xs for the precue repli-

cations; 39.81 and 43.92 Xs for the postcue replications), its higher level in the postcue condition was not due simply to greater practice. In the postcue condition, the lower six stimuli were overestimated, and the upper six stimuli were underestimated (see Figure 3). Minturn and Reese (1951) found overestimation when less than 20 (but more than 5) elements were presented, but the present results indicate that overestimation occurs only under special (e.g., postcue) circumstances.

Individual exponents did not deviate significantly from the normal distribution (dotted lines in Figure 4) in either the precue condition ($D = .120$, $p < .15$) or the postcue condition ($D = .031$, $p < .20$). Since all but 4 of the 100 subjects underestimated the cue stimulus sheet, one might have expected the feedback to prompt nearly all subjects to maintain or increase the exponent. However, the main effect instead was to reduce variability on the exponent, increasing the ex-

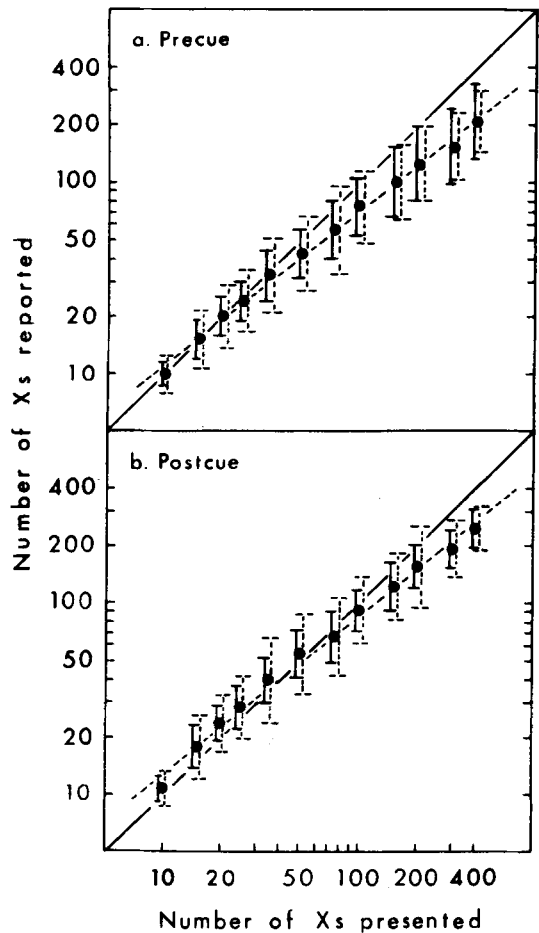


Figure 3. Magnitude estimation: Geometric mean number of Xs reported (filled circles) by number of Xs presented, on log-log coordinates, for precue and postcue conditions. The solid diagonal line (veridical function) has a slope of 1.0. The solid vertical line represents the group standard deviation, and the dotted vertical line represents the mean intrasubject standard deviation across the two replications.

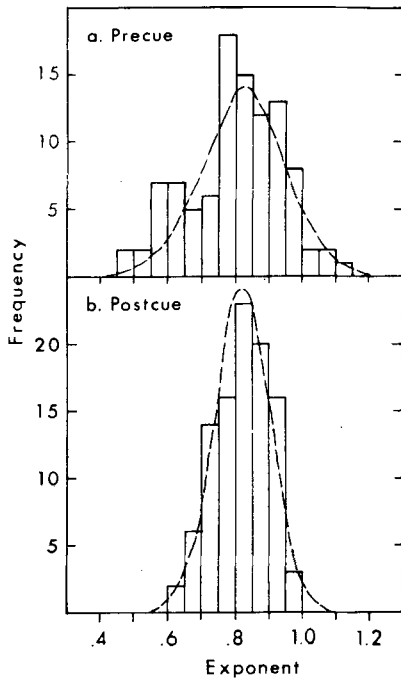


Figure 4. Magnitude estimation: Distribution of exponents for individual subjects. The dotted line depicts a normal distribution.

ponent for subjects in the lower range, but decreasing it for those in the upper range (see Figure 4). Greater practice alone did not reduce the variability in the postcue condition, because variability was stable within each condition (SDs of the exponent were .148 and .143 for the precue replications and .088 and .090 for the postcue replications). The SDs within each condition did not differ significantly ($t < 1$ in all cases), whereas those compared between condition did ($p < .001$ in all cases).

Feedback did not eliminate all intersubject variability, though, because the postcue replications correlated moderately on both the scale factor (+.54) and the exponent (+.61). These correlations are nearly as large as the corresponding ones in the precue condition (+.69 and +.79). Furthermore, the precue and postcue conditions correlated on the scale factor (+.49) and on the exponent (+.50), which again indicates that feedback did not eliminate all consistent intersubject differences. (In every case, $df = 98$ and $p < .001$.)

If only intrasubject variability was present, then the group SDs at particular stimulus sizes (solid vertical lines in Figure 3) would be only 71% ($1/\sqrt{2}$) of the intrasubject SDs (dotted vertical lines in Figure 3). This was approximately true in the postcue condition (65%), but not in the precue condition (81%), in which greater intersubject variability would be expected.

As the number of Xs present increased, the group SDs generally increased in the precue condition (slope of .08), but not in the postcue condition (slope of

.01). Krueger (1982) found a slope of .20, which would have predicted a slope of about .14, given the use of two trials per subject in the present case ($1/\sqrt{2}$ factor). The present subjects, unlike Krueger's (1982), saw more than one stimulus, and they were specially recruited for the numerosity tasks, which might well explain the smaller precue slope (.08) found here.

The intrasubject SDs (dotted vertical lines in Figure 3) did not become progressively larger, but varied nonmonotonically with number of Xs presented. In both the precue and postcue conditions, the intrasubject SDs at the two stimulus extremes (i.e., lowest and highest number of Xs presented) were significantly smaller ($p < .05$, Newman-Keuls tests) than the peak SD value in the middle of the stimulus range. Thus, variability in the magnitude estimation exponent was inter- rather than intrasubject. Eisler (1963) likewise found a nonmonotonic function for the intrasubject SD in the magnitude estimation of length, with a rise in SD in the main portion of the stimulus series being followed by a downturn at the upper end of the series (see his Figure 3).

Discrimination

The mean and SD of the percentage of errors was $31.71\% \pm 7.44\%$. Errors decreased significantly from

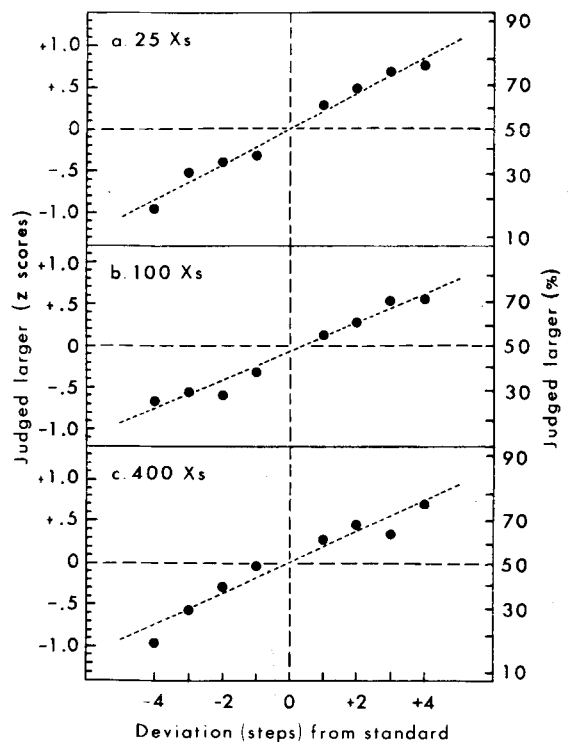


Figure 5. Discrimination: Proportion of comparison stimuli judged larger than standard stimuli (scaled as z scores on left side and as percent scores on right side), by size of standard and size of deviation of comparison stimulus from standard. Step sizes were 1, 3, and 9, respectively, for standards of 25, 100, and 400. Thus, the range of deviations in number of Xs was -4 to +4, -12 to +12, and -36 to +36, respectively.

the first half of the task to the second half, 33.44% vs. 29.98% ($p < .001$). As the comparison stimulus increased in size, the tendency to judge it as larger than the standard increased in an S-shaped manner (normal ogive), producing approximately linear functions when z scores were plotted (see Figure 5). The jnd or differential thresholds (i.e., half the stimulus distance between the 25% and 75% scores) were 3.1, 11.7, and 32.3, respectively, for standards of 25, 100, and 400 Xs. The corresponding Weber fractions were .126, .117, and .081. Similar Weber fractions were found in previous studies. Van Oeffelen and Vos (1982), using standards of 8 to 30, obtained a Weber fraction of .162, whereas Burgess and Barlow (1983) obtained Weber fractions of .100 and .077, respectively, for standards of 25 and 100, with a well-practiced subject (see their Figure 4).

Subjects were only moderately consistent across trials in their sensitivity to differences in numerosity. Percentage of errors in the first half of the task correlated significantly with that in the second half [$r(98) = +.48$, $p < .001$]. Between standards of 25 and 100, the correlation was $+ .28$ ($p < .01$), between 25 and 400, it was $+ .49$ ($p < .001$), and between 100 and 400, $+ .40$ ($p < .001$).

Comparisons Between Tasks

Magnitude production correlated significantly with precue magnitude estimation [scale factor, $r(98) = +.32$, $p < .01$; exponent, $r(98) = +.39$, $p < .001$], as well as with postcue magnitude estimation [scale factor, $r(98) = +.20$, $p < .05$; exponent, $r(98) = +.29$, $p < .01$]. A correlation of $+ .30$ between magnitude estimation and magnitude production exponents was obtained by both Wanschura and Dawson (1974) and Krueger (1972), whose data were reanalyzed by Krueger (1982). The consistent positive correlation on the exponent rules out the possibility that sizeable and consistent individual differences in range of judgments could account for the regression effect (Stevens, 1971), since that would have produced a negative correlation instead (subjects with a much smaller exponent on magnitude estimation because of a constricted range of judgments would tend to have a much higher exponent on magnitude production).

Discrimination errors did not correlate well with any measure, including those involving variability, from the magnitude production or magnitude estimation tasks. Negative correlations were found between discrimination errors and the magnitude production exponent [$r(98) = -.09$] and the precue magnitude estimation exponent [$r(98) = -.15$]. Both correlations were insignificant ($p > .10$ in both cases), as was also the positive correlation on postcue magnitude estimation [$r(98) = +.01$]. Reliability, however, was only moderate within tasks (i.e., between replications). When this was taken into account by applying a correction for attenuation, $r_{xy}/\sqrt{r_{xx}r_{yy}}$, the latter three correlations increased to $-.18$ ($p < .10$), $-.24$

($p < .02$), and $+.02$ (n.s.), respectively. Even so, the intertask correlations are quite low, and they indicate that individual differences in discrimination ability can explain, at best, only about 6% of the variance on the size of the exponent for numerosity. Graf et al. (1974) similarly found small and nonsignificant, but consistently negative ($r = -.01$ to $-.25$), correlations between the Weber fraction and the exponent for brightness under various conditions. They found better evidence across conditions for the theory linking the Weber fraction and the exponent, however, which suggests that the theory is correct but does not hold at the level of individual differences. In the present case, though, it may simply be that numerosity depends on higher level factors. Given that the Weber fraction on numerosity typically is quite large, about .10 (present experiment; Burgess & Barlow, 1983; van Oeffelen & Vos, 1982), an exponent of only about .30 would have been predicted (see R. Teghtsoonian's, 1971, Table 1). People may learn to compensate in part for the tendency to underestimate numerosity, and thereby achieve an exponent closer to 1.0 than the predicted .30.

Sex Differences

Although Krueger (1982) found a higher, more veridical magnitude estimation exponent on numerosity for males (.87) than for females (.78), no significant sex differences were obtained in the present study. Males and females generally had very similar means and SDs, and performance overall was slightly more, not less, veridical for females.

CONCLUSIONS

The present findings indicate that perceived numerosity depends more on higher level cognitive factors (i.e., those involved in interpreting or assigning a particular number to the percept) than on lower level perceptual or sensory factors. In the present case, unlike Krueger's (1982), male and female subjects could self-select their participation on the basis of their interest in the numerosity task, and perhaps as a result there was no tendency for the underestimation to be any greater for females than for males. A single feedback trial had a considerable impact, greatly reducing the variability on individual exponents (Figure 4), while doubling the crossover point between over- and underestimation from 20 to 40 (Figure 3). Performance on the discrimination task did not correlate highly with measures of perceived numerosity, suggesting that perceptual and sensory factors do not play a large part in numerosity judgments.

Consistent individual differences were quite evident in the present study. Individual exponents were moderately correlated between replications in the magnitude production and magnitude estimation tasks, and the nonmonotonic functions obtained on intrasubject SD (dotted lines in Figures 1 and 3) by

stimulus size indicate that the underlying exponent for a given subject did not change across replications. Intrasubject variability apparently involved only the scale factor or the process of retrieving a particular response, and not the exponent. If a longer time interval, say 1 week, had separated the two replications or if the modulus had been changed between them, the correlation might have been largely eliminated (M. Teghtsoonian & R. Teghtsoonian, 1983). The feedback trial, which might be considered to be a change in the modulus, significantly reduced the correlation between replications (from $+0.79$ to $+0.61$, $p < .02$), but even so the postcue exponent correlated $+0.50$ with the precue exponent and $+0.29$ with the magnitude production exponent. These correlations are all the more impressive in view of the sharp decrease in the variability of exponents in the postcue condition (Figure 4).

REFERENCES

- AUERBACH, C. (1971). Interdependence of Stevens' exponents and discriminability measures. *Psychological Review*, **78**, 556.
- BAIRD, J. C. (1970). A cognitive theory of psychophysics: II. Fechner's law and Stevens' law. *Scandinavian Journal of Psychology*, **11**, 89-102.
- BAIRD, J. C., ROMER, D., & STEIN, T. (1970). Test of a cognitive theory of psychophysics: Size discrimination. *Perceptual & Motor Skills*, **30**, 495-501.
- BURGESS, A., & BARLOW, H. B. (1983). The precision of numerosity discrimination in arrays of random dots. *Vision Research*, **23**, 811-820.
- EISLER, H. (1963). Magnitude scales, category scales, and Fechnerian integration. *Psychological Review*, **70**, 243-253.
- GRAF, V., BAIRD, J. C., & GLESMAN, G. (1974). An empirical test of two psychophysical models. *Acta Psychologica*, **38**, 59-72.
- HEINEMANN, E. G. (1978). Discriminability and ratio scaling. In J. C. Armington, J. Krauskopf, & B. R. Wooten (Eds.), *Visual psychophysics and physiology: A volume dedicated to Lorrin Riggs* (pp. 157-166). New York: Academic Press.
- INDOW, T., & IDA, M. (1977). Scaling of dot numerosity. *Perception & Psychophysics*, **22**, 265-276.
- KAUFMAN, E. L., LORD, M. W., REESE, T. W., & VOLKMAN, J. (1949). The discrimination of visual number. *American Journal of Psychology*, **62**, 498-525.
- KRUEGER, L. E. (1972). Perceived numerosity. *Perception & Psychophysics*, **11**, 5-9.
- KRUEGER, L. E. (1982). Single judgments of numerosity. *Perception & Psychophysics*, **31**, 175-182.
- MINTURN, A. L., & REESE, T. W. (1951). The effect of differential reinforcement on the discrimination of visual number. *Journal of Psychology*, **31**, 201-231.
- POULTON, E. C. (1979). Models for biases in judging sensory magnitudes. *Psychological Bulletin*, **86**, 777-803.
- SIEGEL, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.
- STEVENS, S. S. (1957). On the psychophysical law. *Psychological Review*, **64**, 153-181.
- STEVENS, S. S. (1966). Regression effect in psychophysical judgment. *Perception & Psychophysics*, **1**, 439-446.
- STEVENS, S. S. (1971). Issues in psychophysical measurement. *Psychological Review*, **78**, 426-450.
- TEGHTSOONIAN, M. (1965). The judgment of size. *American Journal of Psychology*, **78**, 392-402.
- TEGHTSOONIAN, M., & TEGHTSOONIAN, R. (1983). Consistency of individual exponents in cross-modal matching. *Perception & Psychophysics*, **33**, 203-214.
- TEGHTSOONIAN, R. (1971). On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, **78**, 71-80.
- TEGHTSOONIAN, R., & TEGHTSOONIAN, M. (1978). Range and regression effects in magnitude scaling. *Perception & Psychophysics*, **24**, 305-314.
- VAN OEFFELEN, M. P., & VOS, P. G. (1982). A probabilistic model for the discrimination of visual number. *Perception & Psychophysics*, **32**, 163-170.
- WANSCHURA, R. G., & DAWSON, W. E. (1974). Regression effect and individual power functions over sessions. *Journal of Experimental Psychology*, **102**, 806-812.

(Manuscript received September 14, 1983;
revision accepted for publication April 2, 1984.)