Error patterns in the judgment and production of numerical proportions

J. B. BROOKE and A. W. MacRAE

University of Birmingham, Birmingham, B15 2TT, England

Sinusoidal constant error curves frequently occur in the judgment of numerical proportions, but there are puzzling discrepancies in the direction of constant error reported in various experiments. An experiment showed that the direction of constant error was reversed when the required response changed from judgment to production of proportions. It is concluded that errors are perceptual in origin rather than reflecting response biases. One subject showed an error pattern opposite to the group mean, both in judgment and in production. It is speculated that the discrepancies between reported results may primarily reflect the random assortment of subjects with consistent, but fundamentally different, characteristics.

In experiments on the judgment of visual proportion, subjects are shown patterns made up of elements of two kinds and have to say how many, or what proportions, there are of each kind. Elements in the patterns may be presented simultaneously or successively. Peterson and Beach (1967), reviewing a number of such experiments, report that the constant errors of judgment form roughly sinusoidal curves when plotted ordinally against the proportions to be judged. Those that show overestimation, and hence positive constant error (CE) in the lower half of the scale and negative CE in the upper half, may be approximated by one cycle of a positive sine curve. Those that show the converse pattern of CE may be described as negative sine curves. Positive and negative sine curves correspond to the terms "inverse ogival" and "ogival," respectively, used by Peterson and Beach. The former usage is preferred here because it accords better with our data presentation.

Sinusoidal curves occur both for sequentially presented proportions and for simultaneously presented proportions. However, Peterson and Beach found puzzling discrepancies in the types of curves reported, and could not attribute them to any particular features of the experimental procedures used. Thus, for instance, with sequential stimuli, Pitz (1965, 1966) and Simpson and Voss (1961) reported negative sine CE curves, but Erlick (1964) found positive ones. With simultaneously

presented stimuli, Stevens and Galanter (1957) reported a positive sine curve, and Shuford (1961) a negative one.

In attempting a reconciliation of these apparently contradictory results, it would be useful to know whether the departure from linearity of the function relating judged to actual proportion is caused by perceptual or response factors. It is possible to evaluate their relative contributions by using two procedures—a judgmental response mode in which the subject is shown a stimulus representing a certain value on the stimulus range and is asked to judge its position on the range, and a production response mode in which the subject is asked to adjust the stimulus to a given value. If error is perceptual, a comparison of judgmental and production responses will reveal one pattern of outcomes, whereas if error is attributable to response biases, a different pattern will emerge.

The relation between response magnitude and stimulus magnitude can be examined by determining the functional relation between verbally specified proportion and the proportion of relevant elements in the stimulus display. If no response bias is present, then the relation would be the same whether the subject controls the verbally specified proportion, as in the judgmental response mode, or the proportion in the stimulus display, as in the production response mode. When the relation is the same, then any departure from linearity is attributable to perceptual "error." When the relation is not the same, then the departure from linearity is at least partly attributable to response biases. This is the analysis used by Stevens and Greenbaum (1966) to demonstrate the presence of a response bias in crossmodal matching.

Alternatively, departures from linearity of the function relating response magnitude to stimulus

This work was supported by a research training award to J. B. Brooke by the Science Research Council (U.K.). Requests for reprints should be sent to Dr. J. B. Brooke, Department of Applied Psychology, UWIST, Penylan, Cardiff, CF3 7UX, Wales, or to Dr. A. W. MacRae, Department of Psychology, University of Birmingham, P. O. Box 363, Birmingham, B15 2TT England.

magnitude may be expressed in terms of error. If, instead of response *magnitude*, response *error* is considered as a function of stimulus magnitude, then the absence of response bias is shown by complementary error patterns for judgment and production. This is the case, for example, when a subject overestimates, by giving $X + i$ as an estimate when the displayed proportion is X . This constitutes a positive constant error, i , but when asked to produce that proportion $X + i$, he selects the value X , and thus shows a negative constant error, $-i$. A similar situation arises if he underestimates (says $X - i$) when the true value is X . The constant error is then negative, but if he is asked to produce the proportion $X - i$, he will select X , giving a positive constant error. Thus, if there is no response bias, the judgment error curve will be the complement of the production error curve. If, on the other hand, not perceptual factors but only response biases operate to produce constant error, the constant error curves will be similar in nature for judgment and production (to the extent that the same response biases operate when the subject adjusts proportions as when he specifies numbers). An example of a response bias would be a tendency to avoid responses (whether judgments or stimulus settings) which are near the limits of the possible range.

It is important to notice that when the relation between response *magnitude* and stimulus magnitude is examined, the physical dimension (e.g., the proportion of one type of element in a display) is usually represented on the x-axis of a graph, although it is the independent variable in the judgment procedure but the dependent variable in the production procedure. But when the relation between response *error* and stimulus magnitude is examined, it is more usual to represent the dependent variable, response error, on the y-axis, whether the error is in numerically estimated proportion or in the subjects' settings of proportion in the stimulus display. Since there is no uniformity among authors in the presentation of data and the two conventions require diametrically opposite interpretation, care must be taken in comparing results from different authors.

Although discrepant error patterns have been reported for judgments of proportion, as noted earlier, these are not attributable to differences in the representation of the data. Nor can they be attributed to differences in response mode, since all the experiments required only judgments by the subjects. In this paper, we present results from the scaling of visual proportion using both a judgment and a production procedure. Our concern is with patterns of error, and the constant error curves from the two procedures will be compared to discover whether errors in judgments of proportion are perceptual in origin.

METHOD

Subjects

All subjects were undergraduates of the University of Birmingham who had volunteered for the experiment without payment. Five were male and two were female.

Procedure

In the judgment condition, stimuli consisted of 100 elements arranged in a 10 by 10 matrix, about 75 mm square. Each element was a short horizontal or vertical bar, 5 mm long. Stimuli were generated by a PDP-11/20 computer in the following manner. The number of horizontal bars required in the stimulus was read from a data file, and these bars were allocated to matrix locations chosen at random. The remainder of the 100 matrix locations were then assigned vertical bars. Once the matrix had been filled, the horizontal and vertical bars were drawn on a Tektronix 611 visual display unit in the selected positions. Subjects viewed the matrix by closed circuit TV in an experimental cubicle.

Each trial was preceded by the display of a warning signal. This was the word "ready" presented for 1.5 sec on the screen in capital letters, 20 mm high. A matrix was then displayed on the screen for 7 sec, the first 2 sec being the time required to build up the display on the screen. It was then erased and replaced by a two-digit number (each digit, 20 mm high). The subject could adjust the number to any integer value between 01 and 99 by using a 10-turn potentiometer on the response console. This controlled a voltage that was read by the computer and interpreted as a two-digit number. The subject adjusted the number to what he judged to be the percentage of horizontal bars in the matrix, and then pressed a signal button on the console. The current value of the response number was then recorded by the computer and a new trial was initiated. The display time of 7 sec was chosen to make counting of the elements impracticable.

Each subject began with a practice session of 20 trials and then performed the experimental session in blocks of nine trials each. Before each block, the subject was allowed to rest for as long as he liked. He then initiated the next block of trials by pressing the signal button. The same data file controlled the sequence of stimulus values presented to all subjects, but each subject's sequence began at a different block in the file. Thus each block occurred early for some subjects and late for others.

Two subjects (1 and 2) each made 180 judgments, followed by 180 productions. This proved to be extremely time consuming, and both subjects reported that their attention for the task reached a low level. The other subjects therefore had only 10 blocks of each condition (thus yielding 90 judgments and 90 productions each).

In the production condition, the warning signal was displayed for 1.5 sec, followed by 2 sec of a number, read from a data file and representing the proportion of horizontal bars to which the subject had to adjust the matrix. Next, a matrix with a random number of horizontal bars was drawn on the screen, and when the display was complete, a signal lamp was illuminated. This indicated the start of the period in which the subject could adjust the matrix. By rotating the potentiometer on the response console, vertical elements of the matrix could be transformed into horizontal ones, or vice versa. These transformations occurred in a random manner; that is, if the potentiometer was turned first one way and then the other, the elements it transformed would not as a rule be the same ones. About 2.5 turns of the potentiometer were sufficient to transform the matrix from 99% horizontal to 99% vertical elements.

The subjects were allowed a maximum of 20 sec in which to adjust the matrix to the required proportion of horizontal bars. If, however, they were satisfied that they had adjusted the matrix to the correct proportion, they could stop the trial sooner by pressing the signal button. The number of horizontal bars in the matrix was then recorded by the computer and a new trial was

initiated. In other respects, the procedure was the same as in the judgment condition.

Design

All subjects performed the judgment condition first and then the production condition, with a rest period of about 5 min between the conditions. The reason for presenting the two conditions in this order is that only proportions that were multiples of 10% were presented for judgment, or asked for in the production condition. If some subjects had performed the production condition first, they would have known that only simple decile proportions were being asked for, and this might have influenced their performance on the judgment condition. This problem was avoided by always having the judgment condition first, so the subjects did not learn that the stimuli were always decile proportions.

RESULTS

The CE curves for individual subjects are shown in Figure 1. Trials 1-90 and 91-180 for Subjects 1 and 2 are plotted separately as 1a, 1b, 2a, and 2b

Figure 1. Constant errors in the judgment and production conditions. Each plotted point in an individual curve is the mean error in 10 trials. The nine points in each graph represent errors in judging or producing 10%, 20%, . . . 90% horizontal elements. Errors above the line indicate judgment or production of too large a percentage of horizontal elements. Errors below the line indicate too small a percentage. The horizontal axes are spaced apart by 20 percentage units of error. Subjects 1 and 2 performed both conditions twice (1a, 1b, 2a, and 2b). Only their first performances are included in the mean curves.

respectively. Subjects 1 and 2 show a larger range of errors than the others, but these seem to be real individual differences, for both performed fairly consistently on the first 90 and the second 90 trials. Most curves are noticeably sinusoidal and complementary in direction in the two response conditions. Possible exceptions occur only where the errors are very small and do not show clear sinusoidal patterns.

The complementary nature of the judgment and production conditions is shown by the predominantly negative correlations between CE curves summarized in Table 1. All but one of these correlations, and all the significant ones, are negative.

In the majority of cases, judgment gives a negative curve and production a positive one. Subject 5, however, shows the opposite pattern. Nevertheless, this result accords with the others in that production and judgment produce significantly complementary results, as shown in Table 1.¹

The variability in subjects' responses at each presented proportion is represented in Figure 2. Some subjects show maximum variability when 50% of the elements are vertical, while some show signs of a local minimum there. In general, there is little evidence of the kind of central "anchoring" sometimes reported for spatial scales, though it is true that in the latter case this tends to show up more in group data than with individual results, as indeed occurs here if data are pooled.²

Some subjects, when they had completed the judgment conditions, said that they had made some mistakes with the extreme proportions because they had estimated the proportion of the wrong type of element. These errors were made only in the judgment condition and occurred only for extreme values, as far as could be ascertained. On eight occasions, large percentages of horizontal bars were responded to as being small (i.e., a response of about 10 might be made where one of 90 was correct), and twice the reverse happened. These gross errors are anomalous, since estimation at 10% and 90% was otherwise nearly perfect. Indeed, at these extreme values, the subjects may have had adequate time to count the less numerous type of stimulus element. Because subjects considered that they had made a mistake on these occasions, their recorded judgments were adjusted by subtracting the response from 100.

DISCUSSION

Shuford (1961) had also noted that such errors occur. He did not associate them with any particular response strategy, though he commented that "the extreme errors were due not to random fluctuations in S's perception of the matrix but rather to reversals in the association of element type and proportion" (p. 435). A plausible interpretation of these errors is that subjects always judge the less numerous element

Table 1
Correlations Between CE Curves: Individual Subjects

Subject	r	Significance Level
1a	-.68	.05
1b	-.90	.01
2a	-.84	.01
2b	-.81	.01
3	+.58	n.s.
4	-.42	n.s.
5	-.82	.01
6	-.64	n.s.
7	-.15	n.s.

type and then subtract the judged value from the total when necessary, though sometimes they forget to do so. The two contrary instances here may represent inappropriate application of the subtraction stage. When observing a matrix such as the one used in this experiment, the smaller proportion invariably stands out in a figure-ground fashion and thus is easier to judge. This was commented on by several of the subjects.

An experiment by Kaufman, Lord, Reese, and Volkman (1949) supports this interpretation. Sub-

jects were required to estimate the number of dots on a card, and as the number of dots increased beyond ten they tended to underestimate more and more, their median report time increased, and their median confidence in their estimates decreased. It follows from this that if our subjects had been attempting to judge only the number of horizontal elements in the matrix they should always have underestimated, since they were never fewer than 10 horizontals. In fact, our subjects tended to underestimate only when fewer than half of the elements were horizontal. They tended to overestimate the number of horizontals when these were the most numerous type of element. This accords with the results obtained by Kaufman et al. if we interpret our subjects' strategies as estimation of the smaller proportion (whether horizontals or verticals) followed by a recoding to produce a response in terms of the proportion of horizontal elements, as required by the instructions. Consistent numerical underestimation in their perceptual estimates would then lead to overestimation in those responses arrived at by subtraction, i.e., those where most of the matrix elements were horizontal.

The argument assumes that subjects were not making equivalent underestimates of the total number of elements in the array—if they were, then their judgments of proportion would, if anything, err in the opposite direction, since larger numbers were underestimated proportionately more than small ones. However, this assumption is very plausible. The horizontal elements were variable in number and random in location, as in the Kaufman et al. experiment. The total array was constant in number and arranged in a regular 10 by 10 grid, making it easy to estimate.

The advantage of this strategy for a subject is that it avoids the need for direct perceptual judgments of very large numbers of elements—those which, according to Kaufman et al., would be judged most slowly, with least confidence, and with greatest error.

The results from Subject 5 cannot be accounted for in this way, but Kaufman et al. report only pooled data which may conceal such individual differences. The group results in the present experiment certainly do so. It is an intriguing speculation that the discrepancies between reported experimental curves may, in fact, result from the chance assortment of subjects who follow consistent but fundamentally different patterns, and whose results have nevertheless been pooled.

It is clear from this experiment that, in the case of proportions, the source of constant error is perceptual rather than being a response bias. If subjects preferred to respond near the ends of a scale, for instance, we would expect to find the same sort of CE curve in both response conditions, for the reasons discussed in the introduction. This did not occur

Figure 2. Variable errors in the judgment and production conditions. Each point represents the standard deviation of judgments about the constant error indicated by the corresponding point in Figure 1. The horizontal axes are spaced vertically by 20 percentage units.

with any subject, so we conclude that the errors in judging a scale of proportion are the result of perceptual, and not of response, factors.

REFERENCES

- ERLICK, D. E. Absolute judgments of discrete quantities randomly distributed over time. *Journal of Experimental Psychology*, 1964, **67**, 475-482.
- KAUFMAN, E. L., LORD, M. W., REESE, T. W., & VOLKMANN, J. The discrimination of visual number. *American Journal of Psychology*, 1949, **60**, 498-525.
- PETERSON, C. R., & BEACH, L. R. Man as an intuitive statistician. *Psychological Bulletin*, 1967, **68**, 29-46.
- PITZ, G. F. Response variables in the estimation of relative frequency. *Perceptual and Motor Skills*, 1965, **21**, 867-873.
- PITZ, G. F. The sequential judgment of proportion. *Psychonomic Science*, 1966, **4**, 397-398.
- SHUFORD, E. H. Percentage estimation of proportion as a function of element type, exposure time, and task. *Journal of Experimental Psychology*, 1961, **61**, 430-436.
- SIMPSON, W., & VOSS, J. F. Psychophysical judgments of probabilistic stimulus sequences. *Journal of Experimental Psychology*, 1961, **62**, 416-422.
- STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a variety of perceptual continua. *Journal of Experimental Psychology*, 1957, **54**, 377-411.
- STEVENS, S. S., & GREENBAUM, H. B. Regression effect in psychophysical judgment. *Perception & Psychophysics*, 1966, **1**, 439-446.

NOTES

1. A more complete analysis, taking account of the variability of the subject's estimates, is an analysis of variance of Subject 5's results, in which scale positions are represented as two fixed factors, Scale Half and Distance from Center. Omitting the central (50%) scale value, the interaction of Scale Half by Response Mode is highly significant, $F(1,144) = 17.92$, $p < .001$.
2. The mean curves in Figure 2 are not derived from pooled data but are the means of the relevant curves above them.

(Received for publication January 19, 1976;
revision accepted December 23, 1976.)