The subjective uncertainty of self-estimates

SERGIO CESARE MASIN
University of Padua, Padua, Italy

Self-estimates obtained by the rating method or by the magnitude-estimation method are used unprocessed as scores to test statistical hypotheses in factorial designs. Properly, ANOVAs are based on the assumption that the distribution of scores is normal. In Experiments 1 and 2, observers were asked, after they had produced a given numerical-rating response, n, to evaluate the subjective probability or degree of certainty that other numbers within the scale might be chosen as a rating. It was found that the degree of certainty was asymmetrical with respect to an n, when n was near the end values of the scale. In Experiment 3, a similar result was found for the magnitude-estimation method. The asymmetry indicates that the population density function of the distribution of numerical responses is skewed. It is concluded that the results of ANOVAs should be interpreted with caution.

In an increasing number of psychological reports, researchers report self-estimates obtained by the rating method or by the magnitude-estimation method in factorial designs that test statistical hypotheses by means of the analysis of variance (ANOVA). Properly, ANOVAs are based on the assumption that self-estimates in the various treatment conditions are distributed normally, or at least symmetrically, with equal, or nearly equal, variance (Scheffé, 1959). Symmetry and variance of the population density function should be indicated by corresponding properties of the discriminability within the scale. It seems, therefore, of methodological importance to determine empirically how the discriminability varies within scales.

EXPERIMENT 1

Each observer was asked, after he/she had produced a given rating, n, to evaluate the subjective probability or degree of certainty that other numbers (read by the experimenter), smaller and larger than n, might be chosen by him/her as a rating. It was expected that the parameters of the ensuing certainty curve would be affected by the vicinity of n to the end numerical responses, as a consequence of the change in discriminability along the rating scale.

Method
Observers. The observers were 28 unpaid university students who were recruited as they entered the institute.
Stimuli. In the middle of each of 19 12 × 21 cm rectangular gray backgrounds, all having the same reflectance of .23, a square having a side length of 22 mm was placed. The reflectance of a square was .02, .04, .06, .08, .11, .13, .16, .19, .23, .26, .31, .35, .40, .46, .52, .59, .67, .76, or .87. The squares and the backgrounds were made of gray paper from the set of 19 papers supplied by the NCS Institute (Sweden). The stimuli, viewed from a distance of about 1.5 m, were

shown on the frontoparallel plane by removing a white screen (reflectance of about .90–.95). The illumination level was about 30 lx.
Procedure. The squares were numbered from 1 to 19 in order of decreasing reflectance. Before instructions were given, subjects were shown the odd-numbered squares in the order 1, 3, 5, . . . , 17, 19, so that the observer saw the range of shades. The observer was then given the instructions, and the squares were shown again once in random order. For each square, the observer called out a number from 0 to 100 which represented the lightness of the square. Zero was defined as the whitest white, and 100 as the blackest black ever experienced by the observer. The number 50 had to be assigned to the gray color exactly in the middle of these two extremes. Numbers lower than 50 had to be assigned in proportion to light grays, and numbers larger than 50 had to be assigned in proportion to dark grays. The closer a dark gray was to the blackest black, the larger the number, and the closer a light gray was to the whitest white, the smaller the number.
Observers were also told that after they called out the number as a rating, for some squares, but not for all squares, the experimenter would read some numbers between 0 and 100 from a monitor screen. After having asked the observer whether he/she produced the self-estimates with uncertainty (all observers promptly admitted that the estimates were obviously uncertain), the experimenter instructed the observer to ask himself/herself how certain he/she was that a given number just read from the monitor screen could be assigned to the lightness of the stimulus. The stimulus was always present to the observer until all numbers were read from the monitor screen. The observer rated the subjective probability or degree of certainty using numbers in the range of 0–10. Zero was defined as a 0 probability, meaning that the observer was absolutely certain that the number read from the monitor screen could not be assigned to the stimulus. Ten was defined as the maximum probability or maximum degree of certainty that the number read from the monitor screen could be assigned by the observer. Numbers between 0 and 10 represented intermediate degrees of subjective probability or certainty.
The numbers read from the monitor screen, 15 for each stimulus, are represented on the abscissae of the diagrams in Figure 1, and were read only when the observer called out one of the following numbers as a rating: 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 85, 90, or 95. If an observer called out one of these numbers at least twice, then for that observer, the arithmetic means of the two or more ratings of the subjective probability were taken as single ratings of the subjective probability. A session lasted about 30 min.

Results
The results are reported in the left panels of Figure 1. The mean rated probability, or degree of certainty, is reported on the ordinate of the diagrams. The numbers
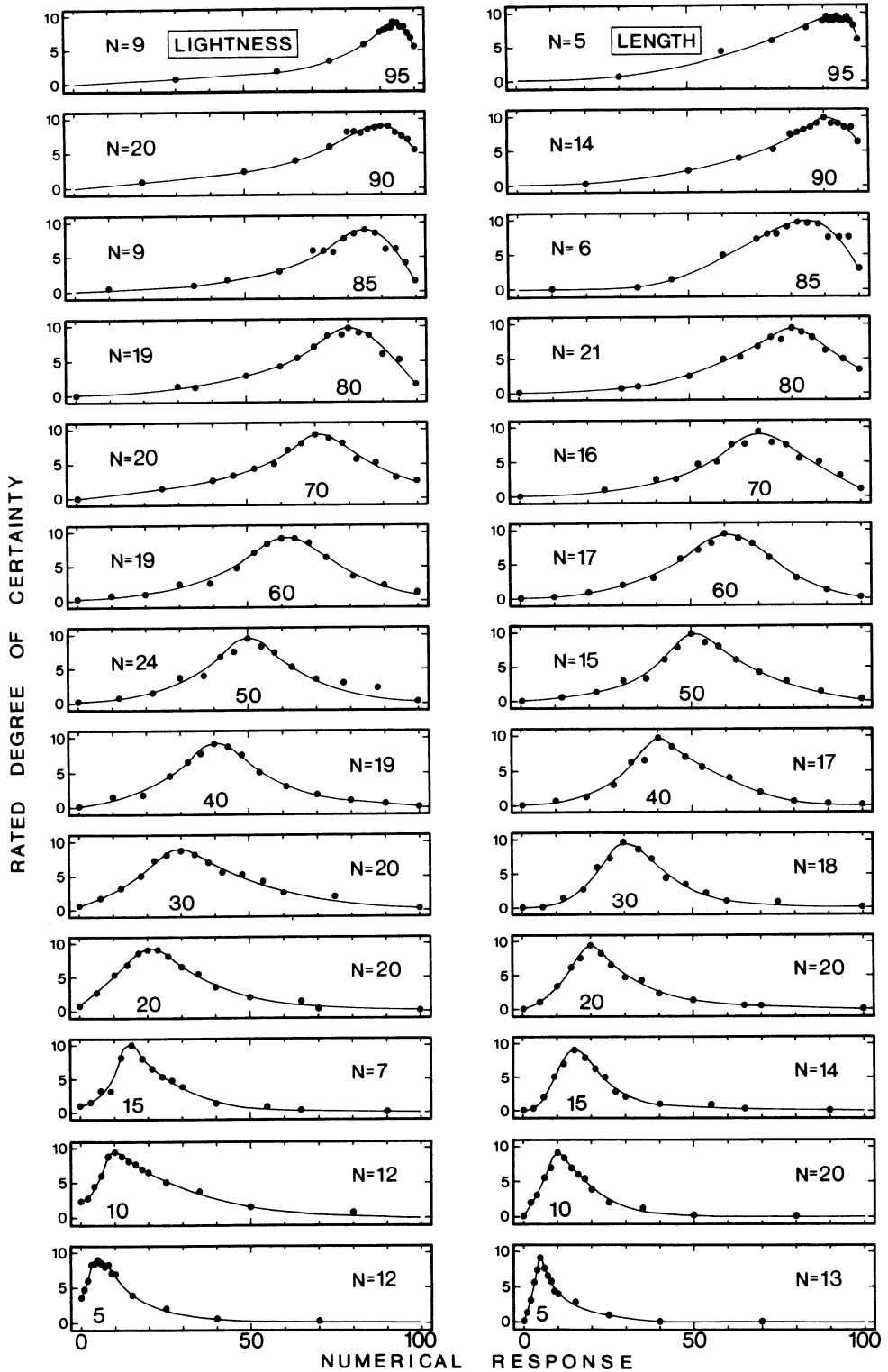
Figure 1. Results of Experiments 1 (left) and 2 (right). The observer's degree of certainty (ordinate) about a given numerical response (abscissa) other than the observer's numerical response (written within a panel).

called out by the observer as ratings and the numbers read by the experimenter to the observer are reported on the abscissae. Each panel refers to a different numerical rating, which is reported on the abscissa of the diagram within the panel. N represents the number of observers (out of 28 observers) who produced the numerical rating indicated in the panel. The standard errors range from about .1 unit to 1 unit of the variable on the ordinate.

As shown in the figure, the certainty curve is symmetrical only for the rating 50. That curve becomes more and more skewed as the rating approaches one of the end numbers of the range 0–100. The skewness may be reasonably interpreted in terms of a decrease in discriminability of the (image or memory) representations of lightness (produced by the numbers read by the experimenter to the observer) toward the middle of the range of the representations of lightnesses.

EXPERIMENT 2

This experiment served to check the repeatability of the above results, using apparent length as a sensory continuum.

Method

Observers. The observers were 26 unpaid university students, recruited as in Experiment 1. None of the students had participated in Experiment 1.

Stimuli. In the middle of 19 12 × 21 cm rectangular white backgrounds, all having the same reflectance of about .90–.95, a horizontal black line was drawn with a length of 5, 10, 15, 20, 25, . . . , 85, 90, or 95 mm. The stimuli were displayed as in Experiment 1.

Procedure. The procedure was the same as that of Experiment 1, except that "square" and "lightness" were replaced by "line" and "length," and except for the following changes. The lines were numbered from 1 to 19 in order of increasing length. The observer had to estimate the length of each line using numbers from 0 to 100. Zero was defined as the nil length, and 100 as the maximum length. It was specified that if the maximum length seemed equal to or less than 100 mm, then the observer could estimate the line lengths in millimeters. The random orders of presentation of the stimuli and of numbers read from the monitor screen were different from those in Experiment 1.

Results

The results are reported in Figure 1, the panels on the right. The diagrams should be interpreted as were those on the left in Figure 1. With minor variations, the results of Experiment 1 are substantially replicated. In both Experiments 1 and 2, the certainty curves have a smaller amplitude near the 0 end of the scale than near 100. Reasonably, this outcome is connected with the observer's special preference for the digits (Baird & Noma, 1978, p. 108).

EXPERIMENT 3

This experiment was undertaken to determine whether the certainty curve was still skewed when the absolute magnitude-estimation method was used.

Method

Observers. There were 21 unpaid university students, who were recruited as in Experiments 1 and 2, but who had not participated in the previous experiments.

Stimuli. The stimuli were pairs of dots on an 18 × 23 cm green-phosphor monitor screen controlled by an Apple II computer. The viewing distance was 90 cm. Each pair was displayed horizontally in the middle of the screen. One dot always appeared in the same position at 3.5 cm from the left side of the screen. The other dot appeared at different distances from the dot on the left. The interdot distances were 19, and could be 2, 4, 6, 8, 10, . . . , 34, 36, or 38 4.7-mm units. The experiment was performed in a dimly lit room, and the dots appeared as luminous tiny points on a dark-gray rectangular screen.

Procedure. The 19 pairs of dots were shown twice in the same random order. The subject was asked to estimate the interdot distance in centimeters (magnitude estimation). Each numerical response was fed into the computer. The computer was programmed to compute the arithmetic mean of the two numerical responses per pair of dots, and then, using these arithmetic means as single numerical responses, to fit the psychophysical function $\ln n = \ln a + b \ln d$ (where n is the numerical response, a and b are parameters, and d is the physical interdot distance) through the data, as soon as the estimation of interdot distances ended. Then, for each observer, the value of n corresponding to the interdot distances, d, equal to 4, 8, 16, and 32 (4.7-mm units) were computed.

Each observer was then given the following instructions. On the upper left corner of the screen, one of four possible target numbers (with one figure after the decimal point) might appear. The target numbers were the four ns computed as above. The target numbers had to be interpreted as the distance in centimeters between pairs of dots. A target number was presented for 5 sec. Then, in the middle of the screen, a pair of dots appeared for 5 sec. The pair of dots was presented exactly as before, except that the interdot distances were those represented on the abscissae of the diagrams in Figure 2. The task of the observer was to rate, on a 0–10 scale, the probability or degree of certainty that the pair of dots that appeared in the middle of the screen was the pair of dots corresponding to the target number in the upper left corner of the screen. As before, 0 was defined as the 0 probability, and 10 as the highest probability, or degree of certainty.
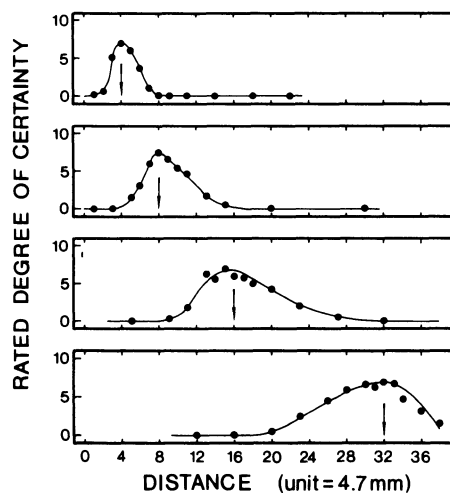
A session lasted about 20 min.

Figure 2. Results of Experiment 3. The observer's degree of certainty (ordinate) in assigning a numerical response in centimeters (previously estimated from the individual psychophysical function, for the interdot distances indicated by the arrows) to a pair of dots whose distance from one another is represented on the abscissa.

Results

The results of Experiment 3 are illustrated in Figure 2. The mean rated degree of certainty is reported on the ordinates, and the interdot distances on the abscissae. The arrows indicate the interdot distances (4, 8, 16, or 32, in 4.7-mm units) from which the target numbers were computed. As may be seen, the certainty curve for the target number nearest to the upper-end numerical response (bottom panel) is also skewed. This indicates, plausibly, that the observer implicitly assumed the existence of an upper sensory bound for apparent distance. This upper bound might have been the size of the monitor screen.

GENERAL DISCUSSION

The marked changes in amplitude of the certainty curves with the distance of the numerical response from the ends of the rating scale seem to be due to the decreased discriminability in the middle of the scale (John, 1969; Montgomery, 1975, 1977). If a certainty curve is truly indicative of the parameters of the population density function of the distribution of numerical reponses, then its marked asymmetry should reflect a skewness in the density function itself. The form of certainty curves indicates that the form of the distribution of numerical ratings approximates that of a beta distribution (Johnson & Kotz, 1970). This conclusion suggests that ANOVAs which are based on the assumption that the density function is normal should be interpreted with caution.

REFERENCES

BAIRD, J. C., & NOMA, E. (1978). *Fundamentals of scaling and psychophysics*. New York: Wiley.

JOHN, I. D. (1969). Stimulus discriminability in the magnitude estimation and category rating of loudness. *Perception & Psychophysics*, **6**, 78-80.

JOHNSON, N. L., & KOTZ, S. (1970). *Distributions in statistics: Continuous univariate distributions* (Vol. 2). New York: Wiley.

MONTGOMERY, H. (1975). Direct estimation: Effect of methodological factors on scale type. *Scandinavian Journal of Psychology*, **16**, 19-25.

MONTGOMERY, H. (1977). Magnitude scales, category scales, and the general psychophysical differential equation. *Perception & Psychophysics*, **21**, 217-226.

SCHEFFÉ, H. (1959). *The analysis of variance*. New York: Wiley.