

Different biases in the methods of constant and single stimuli

SERGIO CESARE MASIN
University of Padua, Padua, Italy

The method of constant stimuli is unreliable, because it causes an unpredictable constant error. An explanation of the constant error in terms of adaptation-level theory implies that the point of subjective equality changes with time. When the method of constant stimuli was used, the constant error was present from the first response and did not change substantially with time, thus indicating that an explanation in terms of an adaptation level is implausible. A modified version of the method of single stimuli was also used, in which observers had to estimate the magnitude of the length of both lines in a pair. The use of a direct method produced a reversal of the direction, and a substantial reduction in magnitude, of the constant error.

The method of constant stimuli (MCS) produces a constant error whose magnitude depends on the way the variable stimuli are presented (Doughty, 1949; Levison & Restle, 1968; Restle & Levison, 1971). Levison and Restle (1968) used a fixed standard line (S), always presented first, followed by one of seven variable lines (Vs). Observers reported whether V was longer or shorter than S. The mean point of subjective equality (PSE) fell between S and the arithmetic mean of the Vs, with the magnitude of the constant error depending on the spacing between the Vs. Levison and Restle (1968) interpreted their findings in terms of adaptation-level theory (Michels & Helson, 1954), whereas Erlebacher and Sekuler (1971; Sekuler & Erlebacher, 1971) contended that the same results could be explained on the basis of a tendency of observers to use (the two) available responses ("shorter" and "longer") with equal frequency (Doughty, 1949; Koester & Schoenfeld, 1946; Preston, 1936).

According to Helson (1947) and Helson, Michels, and Sturgeon (1954), V is judged relative to adaptation level, which is conceived as a function of S and the average of previous Vs. Sekuler and Erlebacher (1971) argued that there would be no plausible reason to consider the S separately from the Vs, and that the adaptation level should affect both S and V in the same amount, leaving the PSE unchanged. According to Helson (1947, p. 4), the reason why S should be considered separately from the Vs is that the Vs, being the objects of judgment, tend to form a group, leaving S "more or less" outside. Woodruff, Jennings, and Rico (1975) showed that the constant error in the MCS occurs also when the observer does not know which stimulus is the S or which is the V. Thus if S contributes separately to the adaptation level for the reason suggested by Helson, then that contribution should

show up only when the observer becomes aware that one of the two compared stimuli (S and V) does not change in successive presentations. It seems plausible then, in order for S to be separated from the Vs, that all Vs forming a pair with S in a series should be presented at least once, or possibly two, three, or more times. Therefore, if the adaptation level depends on a psychological separation of S from the Vs, it may be predicted that the adaptation level for the average observer changes gradually (since different observers would become aware of S at different times, from the beginning of presentation of stimuli) or, at least, changes after all the Vs in a series are presented once.

If observers have the tendency to use the responses "shorter" and "longer" with equal frequency, it may be that (1) for the response tendency to be activated, the observer must first become aware of the preponderance of one of the two responses, or (2) the response tendency is activated immediately. In the first event, a certain number of responses must initially be required. Therefore, it seems reasonable to predict that, on average, the response tendency would take effect gradually (as observers would become aware of a preponderance of one response at different times), or at least after all the Vs in a series are presented once. By comparison, in the case of an immediate activation, the response tendency should be exhibited in the observer's first few responses.

If it is true that the constant error produced by the MCS depends either on adaptation level or on an awareness of a preponderant response, then it may reasonably be expected that the constant error is nil, or at least substantially reduced, when the observer gives only one response per V. The following experiment was designed to test this prediction.

EXPERIMENT 1

Method

Observers. The observers were 100 university students and members of the psychology department at the University of Padua.

I am now at the University of Udine, Udine, Italy. I wish to thank M. Teghtsoonian and G. Vicario for comments, and S. Bettella for help in cross-checking the microcomputer programs. Mail may be directed to me at Institute of Philosophy, Pedagogics, and Teaching of Modern Languages, Via Antonini 8, 33100 Udine, Italy.

Stimuli. The stimuli were produced by an Apple II computer in the middle of a green-phosphor 18×23 cm monitor screen. The stimuli were green vertical lines having a luminance of 7 cd/m². The luminance of the dark-gray background was .6 cd/m². The method of constant stimuli was used, with S always presented first and V second. The Vs (V1-V5) were 77, 79, 81, 83, or 85 pixels, corresponding to a length on the monitor screen of 39.72, 40.75, 41.78, 42.81, or 43.84 mm, respectively. The viewing distance was 125 cm. Both S and V lasted 1.5 sec, with an interstimulus interval of 5 sec. A brief acoustic signal preceded both S and V by 1.5 sec. A question mark appeared in the left-hand bottom corner of the screen 1.5 sec after the V disappeared. After the question mark appeared, the observer responded "V shorter than S" or "V longer than S." Then, after 1.5 sec, the acoustic signal was repeated, and 1.5 sec later S appeared. Thus the minimum time interval between V and the succeeding S was 4.5 sec.

Procedure. The observer sat at a small table provided with a chin-rest to keep the eyes level with the stimuli. He/she was told to say whether the second line was longer or shorter than the first line. It was stressed that he/she had to rely on very small impressions of a difference in length. In case he/she did not detect even an infinitesimal difference, he/she had to respond at random "longer" or "shorter," because no "equal" response was allowed.

Five groups of 20 observers each were formed, all having the same instructions and stimulus conditions, except for the length of the S. For one group S=V1, for another S=V2, for another S=V3, for another S=V4, and for the last S=V5. There were 10 series of five trials, in each of which all five Vs (each coupled with S) were presented once in random order, different for each observer.

Results

The results are depicted in Figure 1, which shows the mean PSE for each series. Each panel shows the results for one group of observers. The vertical bars indicate the double standard error, computed by the Thomson-Gadum formulas (see Masin & Vidotto, 1982).

In each panel, on the right, an empty dot represents the grand-mean PSE, which was computed by fitting (using the weighted least squares method) the psychometric function from all responses of all 20 observers. The standard error, represented by a vertical bar, was computed from the sample of 20 individual PSEs of a group (each obtained by fitting—using the weighted least-squares method—the psychometric function from the 10 responses per V of a single observer). The grand-mean PSE for S=V1 is not statistically different from S [$t(19) = 1.53$].

The grand-mean PSEs for S=V2, S=V3, S=V4, and S=V5 are statistically different from the corresponding S, at the .05 level [$t(19) = 2.6, 4.3, 8.7, \text{ and } 9.4$, respectively].

The horizontal line in each panel represents the value of S. In the inset in Figure 1, the constant-error percentage, $P = 100[(PSE - S)/S]$, is represented (abscissa) as a function of the value of S (ordinate). The horizontal bars indicate the double standard error of P, which was obtained by multiplying the standard error of the grand-mean PSEs of a panel by the ratio 100/S. An analysis of regression of individual Ps over Vs showed that the slope of the regression line was statistically different from zero [$t(97) = 6.24, p < .00005$] (an outlier was omitted).

Discussion

The grand-mean PSEs (except that for S=V1) differ from the corresponding S, even with the first responses to the five Vs. Thus it seems unlikely that the constant error, P, produced by the MCS is due either to adaptation level or to a growing awareness of unbalanced response frequencies. Both of these hypothetical constructs rely on time for full development. The results seem to agree with the possibility that the tendency to balance response frequencies is activated automatically, as soon as the observer starts responding.¹

If the observer is allowed the use of more than two response categories, the tendency to equate response frequencies should not exist any longer, or at least should be strongly reduced, simply because there would be many alternative responses from which to choose. Magnitude estimation allows the use of the maximum number of response categories. In Experiment 2 the observer was asked to estimate the magnitude of the apparent lengths of S and V. Whether V was judged shorter or longer than S could then be deduced by comparing the two corresponding numerical responses.

EXPERIMENT 2

Method

Observers. The observers were 60 university students and members of the psychology department, who had not participated in Experiment 1.

Stimuli. The stimuli were the same as in Experiment 1.

Procedure. Three groups, each of 20 observers, were formed. All groups had the same stimulus conditions as in Experiment 1, except that the length of S was V1 for the first group, V3 for the second group, and V5 for the third group. The method of single stimuli (MSS) was used. All three groups were instructed to estimate the length, in centi-

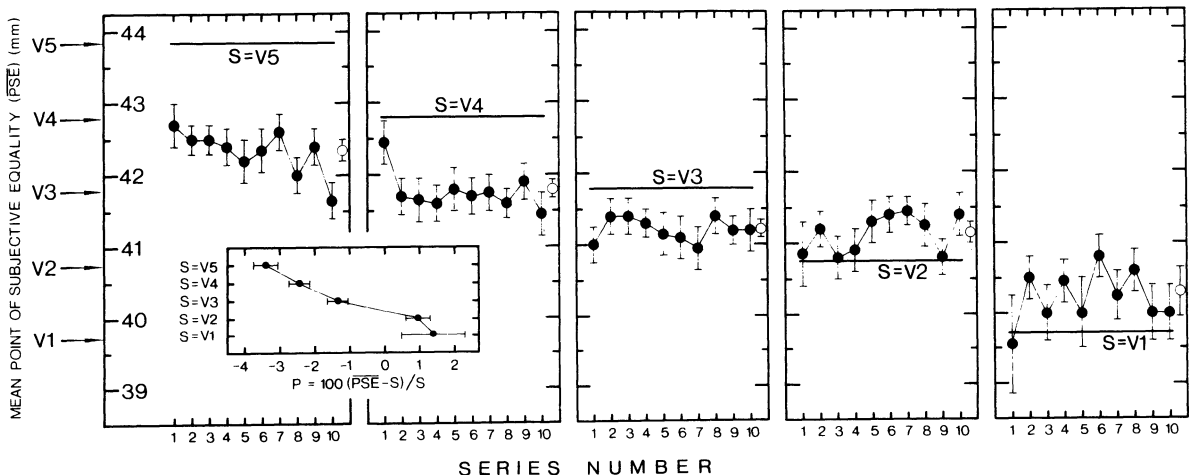


Figure 1. Each panel shows the results for a different group of 20 observers in Experiment 1 (see text for explanation).

meters and millimeters, of a vertical line in the middle of the screen, and then to use that number as a unit of measurement to express the apparent length of subsequent lines. A verbal numerical example was given. Observers were also told that the method did not allow assigning the same number to two lines of a pair of lines, and that, whenever they did so, they had to change the second number by adding or subtracting, at random, 1 mm. The random orders of presentation were different for each observer.

Results

It was assumed that the numerical responses to V and S represented their apparent lengths. The data were processed as in Experiment 1. The results are depicted in Figure 2 (all symbols have the same meaning as in Figure 1) and should be compared with the results depicted in Panels 1, 3, and 5 in Figure 1.

The empty dots, on the left in Figure 2, indicate the grand-mean PSEs, and the corresponding vertical bars indicate the double standard error, both computed as in Experiment 1. None of the grand-mean PSEs were statistically different from the corresponding S [$t(19) = .84, 1.11, \text{ and } .85$, respectively, for $S=V1, S=V3$, and $S=V5$]. The very large standard errors in the condition $S=V1$ are due to the great eccentricity of the psychometric functions (see, e.g., Masin & Vidotto, 1982, Figures 1 and 2). In the inset in Figure 2, the constant-error percentage, P , is represented as a function of S. An analysis of regression of individual P s over V s showed that the slope of the regression line was statistically different from zero [$t(51) = 3.08, p < .005$] (seven outliers were omitted; with the omission, the t test was conservative).

Discussion

The results reported in Figure 2 confirm the experimental hypothesis. The constant error is reduced when a great number of response categories is allowed. It is not abolished completely, as the regression analyses showed that the P s in the insets in Figures 1 and 2 manifest opposite trends. At this stage of research, the reason for these opposite trends remains unknown. It seems plausible, however, that the smaller constant error produced by the MSS is the result of some response bias, since the stimuli in Experiments 1 and 2 were equivalent.

The procedures used in Experiments 1 and 2 differed, however: in Experiment 2 observers responded both to S and to V. Strictly, the results of Experiment 1 should be compared with the results obtained from an experiment in which the responses are applied both to S and to V, but in which the response categories are the same as in Experiment 1. Experiment 3 served to check this procedure.

EXPERIMENT 3

Method

Observers. Twenty university students, who had not participated in Experiments 1 or 2, took part in the experiment.

Stimuli. The stimuli were the same as in Experiments 1 and 2.

Procedure. The standard was V5 for all observers, who were asked to judge whether both S and V were longer or shorter than the preceding line. The sequence of stimuli was the same as in Experiment 1: (S_1, V_1), (S_2, V_2), (S_3, V_3), etc., but the instructions were radically different. The observer judged each line in succession, relative to the preceding line, without overtly designating either line as S or as V. Thus, V_1 was compared with S_1 , S_2 with V_1 , V_2 with S_2 , S_3 with V_2 , and so on, until all 50 pairs were shown. Then, to compensate for the absence of a V_0 for comparison with S, a final S was shown. On average, an S appeared approximately 8–9 sec after a V. The random orders of presentation were different for each observer.

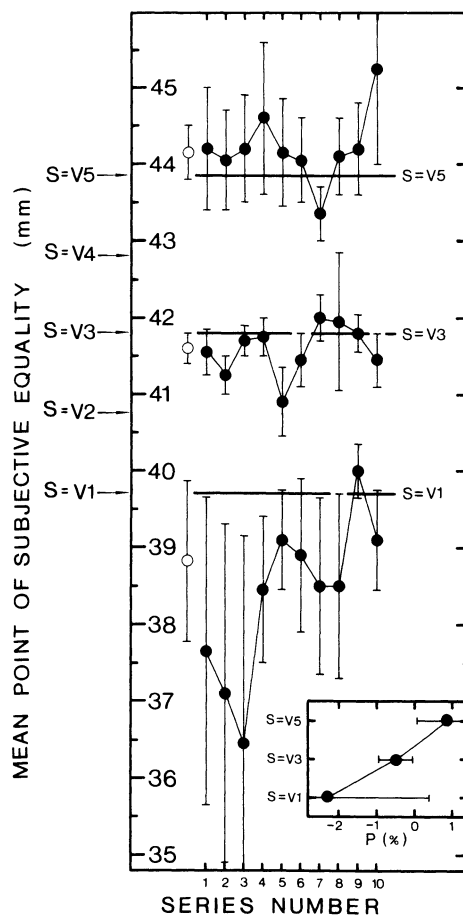


Figure 2. Illustration of the results of Experiment 2 (see text).

Results and Discussion

The results are depicted in Figure 3 (the symbols are the same as those used for Figure 1). The data derived from the explicit responses to the Vs (i.e., from the explicit comparisons of V with the preceding S) were processed exactly as in Experiment 1. The results are illustrated by the bottom graph (indicated by the initials V c.w. S). The vertical bars indicate the double standard error.

Since the pairs of S and V were presented one after the other, the implicit response to V can be deduced from the explicit response given to an S. If S was explicitly judged as shorter than the preceding V, then that V was implicitly judged as longer than S. On the contrary, if S was judged as longer, then the preceding V was judged as shorter. The results of the implicit comparison of V with S (deduced from the explicit comparison of S with V) are illustrated by the upper graph (initials S c.w. V) in Figure 3. The vertical bars indicate the double standard error (to avoid confusion, these bars are sometimes drawn only partially).

The grand-mean PSEs (empty dots), and the corresponding standard errors (vertical bars), computed as in Experiments 1 and 2, are indicated on the right in

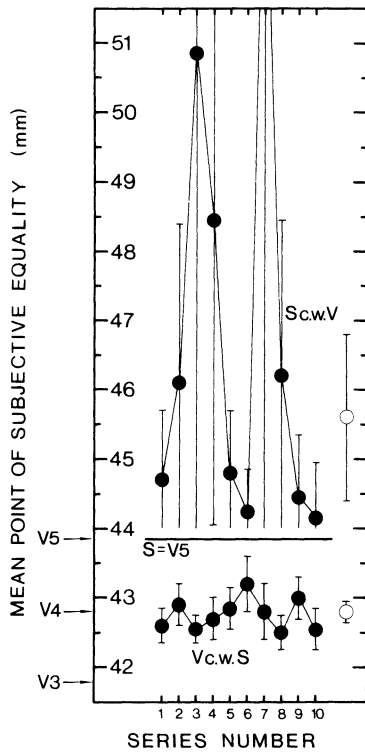


Figure 3. Illustration of the results of Experiment 3 (see text).

Figure 3. The standard error for S c.w. V was computed, excluding 2 observers who produced psychometric functions that were practically flat (due to the very few implicit responses "V longer"). The constant-error percentage, P , for the V c.w. S was -2.42 [its difference from zero was highly significant: $t(19) = 8.2$]. The difference of P from zero for S c.w. V was not significant [$t(17) = 1.43$]. Inspection of the individual results showed that the number of implicit "V is longer" responses barely exceeded the number of explicit "V is longer" responses for only 4 of the 20 observers. Thus, given the great eccentricity of the psychometric function, the difference of $P = 3.99$ from zero should be considered at least as statistically significant as that of $P = -2.42$ from zero. (Experiment 3 was repeated for a parallel study by Masin, Mazzoni, and Vallortigara, 1987. Using 120 observers, the difference of P from zero for S c.w. V was highly significant.)

The result for the V c.w. S should be compared with that reported in the first panel in Figure 1. As may be seen, there is no substantial difference. This shows that, in itself, the judging of both S and V is not the cause of the reduction in magnitude (obtained in Experiment 2) of the constant error. Except for very small statistical fluctuations, the mean PSE for the V c.w. S does not vary substantially with the series number. This result strengthens the conclusions reached in Experiment 1.

For the S c.w. V, the PSE is above V5, in agreement with previous findings (Woodruff et al., 1975). This

presentation-order effect still awaits a satisfactory explanation.

GENERAL DISCUSSION

The results of the experiments reported above strengthen the conclusion that judgment mechanisms in the MCS that involve adaptation level, or a growing awareness of some response tendency, are implausible—at least within the first 50 responses—because the constant error is present from the very first responses to the Vs. Instead, the same results seem to confirm the existence of an immediate response bias in the forced-choice procedure of the MCS.

From the methodological standpoint, the MSS should be preferred, because the constant error it produces seems to be substantially smaller than that produced by the MCS. The MSS should be used jointly with direct methods in the modified version proposed here, where the S is coupled with a V and their magnitudes or differences are self-estimated.

REFERENCES

- DOUGHTY, J. H. (1949). The effect of psychophysical method and context on pitch and loudness functions. *Journal of Experimental Psychology*, **39**, 729-745.
- ERLEBACHER, A., & SEKULER, R. (1971). Response frequency equalization: A bias model for psychophysics. *Perception & Psychophysics*, **9**, 315-320.
- HELSON, H. (1947). Adaptation-level as a frame of reference for prediction of psychophysical data. *American Journal of Psychology*, **60**, 1-29.
- HELSON, H., MICHELS, W. C., & STURGEON, A. (1954). The use of comparative rating scales for the evaluation of psychophysical data. *American Journal of Psychology*, **67**, 321-326.
- KOESTER, T., & SCHOENFELD, W. N. (1946). The effect of context upon judgment of pitch differences. *Journal of Experimental Psychology*, **36**, 417-430.
- LEVISON, M. L., & RESTLE, F. (1968). Invalid results from the method of constant stimuli. *Perception & Psychophysics*, **4**, 121-122.
- MASIN, S. C., MAZZONI, G., & VALLORTIGARA, G. (1987). The first five responses in the method of constant stimuli. *Canadian Journal of Psychology*, **41**, 80-83.
- MASIN, S. C., & VIDOTTO, G. (1982). A review of the formulas for the standard error of a threshold from the method of constant stimuli. *Perception & Psychophysics*, **31**, 585-588.
- MICHELS, W. C., & HELSON, H. (1954). A quantitative theory of time-order effects. *American Journal of Psychology*, **67**, 327-334.
- PRESTON, M. G. (1936). Contrast effects and the psychophysical judgments. *American Journal of Psychology*, **48**, 389-402.
- RESTLE, F., & LEVISON, M. (1971). Method of constant stimuli: Invalidity to the third power. *Perception & Psychophysics*, **9**, 312-314.
- SEKULER, R., & ERLEBACHER, A. (1971). The invalidity of "invalid results from the method of constant stimuli": A common artifact in the methods of psychophysics. *Perception & Psychophysics*, **9**, 309-311.
- WOODRUFF, B., JENNINGS, D. L., & RICO, N. L. (1975). Time error in lifted weights as affected by presentation order and judgment mode. *Perception & Psychophysics*, **18**, 98-104.

NOTE

1. It is interesting to note that, without giving an explanation, Levison and Restle (1968) excluded the first 28 responses of each observer from the data they presented. This may mean (1) that they also thought that a certain number of responses are needed before the adaptation level is changed, or (2) that the first responses by an observer were characterized by some imprecision or greater variability. The present results show that exclusions of data are not necessary.