

The invalidity of "invalid results from the method of constant stimuli": A common artifact in the methods of psychophysics

ROBERT SEKULER* and ALBERT ERLEBACHER
Northwestern University, Evanston, Illinois 60201

A description by Levison and Restle (1968) of deficiencies in the method of constant stimuli is shown to be logically in error. Their data, as well as those from other experiments, are interpretable in terms of a response bias that may distort data in a variety of psychophysical studies. This bias is the tendency of Ss to use available responses with equal frequency.

In a recent methodological paper, Levison and Restle observe that "... the Method of Constant Stimuli may lead an unwary investigator into a wrong conclusion [1968, p. 122]." We heartily agree. The purpose of the present paper is to demonstrate that Levison and Restle themselves were led unwarily into just such an incorrect conclusion by the operation of an artifact that may have widespread influence in most psychophysical experiments, i.e., the tendency of Ss to use available responses with equal frequency.

Among the many choices which precede the execution of any psychophysical experiment are those concerned with the potential stimulus ensemble for the experiment. Many scaling experiments have shown that the scale value for any particular stimulus can be determined, at least partially, by the particular set of other stimuli to which S is exposed (e.g., Parducci, 1965; Poulton, 1968). Levison and Restle (1968) set out to study analogous effects of stimulus ensemble upon data obtained with the method of constant stimuli. In their experiment, four groups of Ss were run, each with a different set of comparison stimuli. The Ss were to judge, on each trial, the length of a standard line (held constant for all groups at 8.25 in.) relative to one of the comparison lines chosen from the stimulus ensemble for their group. The mean length of the comparison lines ranged, between groups, from 6.93 to 9.30 in. The basic finding was that the mean point of subjective equality (PSE) varied systematically among the groups. For example, when the comparison stimuli had a mean length of 9.30, the PSE was 8.35; when the mean of the comparison stimuli was only 6.93, the PSE was reduced to 7.93. Thus, the PSE is shifted away from the point of physical equality (PPF)

toward the arithmetic mean of the comparison series.

Levison and Restle offer an interpretation of their results in terms of adaptation-level theory. Citing Helson (1964), Levison and Restle argue that "S can be thought to rate the comparison stimulus relative to a comparison adaptation level (CAL), and PSE is that stimulus rated at CAL [1968, p. 122]." There may be disagreement on whether adaptation-level theory is a theory of perception (Stevens, 1958). We believe, however, that the theory was intended to be a theory of perception (Helson, 1964), and many papers have used the theory to treat allegedly perceptual phenomena, such as effects of backgrounds or surrounds (Helson, 1959; Restle, 1970) and adapting stimuli (Bell & Bevan, 1968).

Certainly, nonperceptual phenomena can also be described by the mathematics of adaptation-level theory. The fact that the data from some experiment are consistent with the equations of that theory should not imply that perceptual phenomena are in fact responsible for that consistency. An example of such an experiment is that of Levison and Restle (1968). An explanation far simpler than a perceptual one can be made of their data. First, consider the logic of their position. Let us assume that the ensemble of comparison stimuli to which S is exposed actually does alter his perception. Since the method of constant stimuli, as used by Levison and Restle, requires the presentation of both standard and comparison lines on each trial, it is reasonable to expect that any change in the perception of the comparison stimuli will be accompanied by a change in the perception of the standard. (How, for instance, can S tell which is which?) Since the adaptation-level effect should alter the perception of both comparison and standard lines, the PSE should be unaffected by the change in adaptation level. The following example may help to

make this argument more concrete. Imagine two stimulus lines, a standard and a comparison line, of different lengths. In the absence of complications introduced by constant errors, we expect that Ss will judge correctly the order relationship between their lengths. This order judgment, of course, is the basic measurement operation of the method of constant stimuli. From a series of repeated presentations of these stimuli, we can estimate the probability that S will judge one longer than the other. Now let us introduce an experimental manipulation that can be considered equivalent to the perceptual effects of changes in adaptation level. Let us alter the appearance of both standard and comparison lines by imposing some kind of monotonic linear transform on S's perceptual system; e.g., let him view both stimuli through a magnifying glass. Unless the transform is so radical as to take us out of that portion of the perceptual system's dynamic range over which it was approximately linear, the probability of judging the standard longer than the comparison line should be unchanged from the first set of measurements. Since the PSE depends upon the probability that a given standard line will be judged longer than the standard, having S view both standard and comparison lines through a low-power magnifying glass will yield the same PSE as would be obtained with unmagnified viewing. Since the CAL effect postulated by Levison and Restle would produce an equivalent transform of the perception of the standard and each comparison stimulus, it cannot possibly be responsible for the shift in PSE that they obtained.

A more likely explanation can be made in terms of certain well-known response biases that Ss in psychophysical experiments exhibit. There is considerable evidence (Parducci & Haugen, 1967; Tunc, 1964) that when Ss are given a set of responses to use in a psychophysical experiment, other things being equal, they will tend to use all the responses with approximately the same frequency.

How might such an equal-response tendency affect the experiment reported by Levison and Restle? Since most of their Ss had psychometric functions which were monotonic with length of comparison line, it is not possible to argue that Ss gave "longer" and "shorter" judgments in a completely arbitrary fashion, subject only to the constraint of approximately equal frequency. Such a use of the two responses would have produced quite striking psychometric functions of zero slope and indeterminate PSE. It is possible to suggest, however, that on some portion of the trials, when S was less than completely

*Address: Department of Psychology, Northwestern University, Evanston, Illinois 60201.

certain of the correct response, he selected from the response categories in such a way as to maintain an approximate balance between their frequency of use. A response tendency of this form, or some variant of it, would cause the PSE to shift with changes in the composition of the comparison stimulus ensemble. An example, using a somewhat reduced number of comparison stimuli, may help to clarify this argument. Assume a standard line length of 100 mm and only two comparison lengths, 96 and 104 mm. (Since the argument is not affected by this restriction, assume conditions that result in no constant error.) If S used sensory information correctly on 80% of the trials and guessed, subject to response-equalization tendency, on the remaining 20%, the PSE would be at PPE-100 mm. Now let us change the stimulus ensemble, so that the comparison lengths are 96 and 108. Again, let S use sensory information correctly on 80% of the trials, and give equal numbers of "longer" and "shorter" judgments on the remainder. This would produce a PSE at the midpoint of the comparison stimulus series, 102 mm. This simple example shows how a tendency to use the two responses with equal frequency would shift the PSE to the midpoint of the stimulus series.

This simple model is not by itself sufficient to explain fully the findings of Levison and Restle. Recall that they found the PSE to be shifted away from the PPE and toward, but not to, the mean of the comparison series. In order to account for Levison and Restle's data in terms of response bias, we must take both the particular comparison stimulus ensemble members, as well as the likely time order error, into account. For example, it is reasonable to assume that not all comparison stimuli are equally likely to result in that uncertainty of judgment which, in our crude model, leads to the arbitrary selection of that response which is required to generate approximately equal use of the two responses. Certainly, judgments of the length of a comparison stimulus that is very similar to the standard should be more affected by the equal-response tendency than would be judgments of a comparison stimulus that is several jnds away from the standard in length. But regardless of the nuances of the model that would actually be required to "predict" Levison and Restle's data in detail, it should be apparent that a tendency toward equal use of the two response categories would play a key role.

Another example may help to demonstrate further the absurdity of a perceptual explanation in the Levison and Restle experiment and, at the same time, suggest a remedy for the artifacts in

question. In their experiment, the standard and comparison lines were presented sequentially. If the order of the two stimuli varied randomly from trial to trial, it would be impossible for any perceptual change due to adaptation level to differentially affect the two stimuli, since only E knows which is "standard" and which is "comparison"—as indeed is the case when the order is not balanced. However, with this method, any response equalization bias pushing the PSE in one direction on half the trials will be counterbalanced by an equal shift in the opposite direction on the other trials. The net PSE would then remain equal to the PPE.

The perceptual explanation offered by Levison and Restle is impossible on logical grounds, and we have presented a reasonable explanation for their results. We believe, however, that the same explanation may hold for the results of many other psychophysical experiments where the correctness of our hypothesis becomes an empirical matter. As an example, let us take a recent paper by Bell and Bevan (1968). Those investigators report four experiments, which they claim demonstrate effects of various anchor stimuli upon Gestalt principles of perceptual organization, e.g., proximity, closure, similarity, and good continuation. A description of one of their experiments, on proximity, will serve to represent their basic technique, as well as to point up the importance of the equal-frequency principle as a possible determinant of their findings.

The stimuli were 11 matrices of dots, which differed in the relative horizontal and vertical separations between pairs of neighboring dots. When its horizontal separations between dots were larger than its vertical, a matrix seemed to be organized into columns; when its vertical separations were larger than its horizontal separations, a matrix seemed to be organized into rows. In the control condition, the 11 stimuli were presented singly in five blocks to Ss who judged whether each was organized into columns or rows. Calculation of the present "row" responses for each stimulus permitted Bell and Bevan to generate psychometric functions and to estimate the indifference point (the stimulus which would produce 50% "row" and 50% "column" judgments). Two other groups were run in anchor conditions. For one group (the "row" anchor group), the procedure was as above, except that on every other trial Ss were shown and asked to categorize a stimulus that was at the "row" extreme of the continuum. For the other group (the "column" group), the stimulus shown on every other trial was one at the "column"

extreme of the continuum.

The results of Bell and Bevan show that, with the introduction of the column anchor, the 11 members of the original stimulus series were more frequently judged to be organized as rows. Introduction of the row anchor had the opposite effect, causing the frequency of "column" responses to be increased for the 11 members of the original stimulus series. They describe their experiments as supporting Helson's proposition that "there are spatial adaptations, no less than chromatic adaptations, which are responsible for the way spatial patterns are formed [Helson, 1966, p. 173]."

While there have been several experimental demonstrations that spatial adaptation is a major determinant of pattern perception (Blakemore & Sutton, 1969; Held, 1962; Pantle & Sekuler, 1968), we do not think that the experiments of Bell and Bevan constitute such a demonstration. The first reason for our doubts is that in other demonstrations of spatial adaptation, an adapting pattern of a particular spatial frequency (cycles/degree of visual angle) exerts its greatest effect on test patterns of a similar, though not necessarily identical, spatial frequency. In the psychometric function shown by Bell and Bevan (1968, p. 672), however, it seems that the supposed adapting stimuli (i.e., the anchors) have effects that are not systematically related to the difference between the columnarity of the anchor and the columnarity of each of the 11 test stimuli. A second reason for doubting Bell and Bevan's purported demonstration of spatial adaptation is that other studies show that an adapting pattern of one orientation, say vertical bars, has no effect on the threshold for a pattern having an orthogonal orientation, i.e., horizontal bars (Houlihan & Sekuler, 1968; Pantle & Sekuler, 1969). Contrary to these findings, the data of Bell and Bevan show an effect of a row anchor upon the judgment of stimuli having extreme columnarity, and vice versa.

It is very likely that the Bell and Bevan experiment had nothing whatever to do with perceptual adaptation. Their data are easily interpreted in terms of the tendency of Ss to use available responses with equal frequency. Consider, for example, the effects of the row anchor, as interpreted by the equal-frequency principle. The row anchor, an extreme stimulus, would produce "row" responses on almost every one of its appearances. This would reduce the probability that Ss would make "row" responses to the relatively ambiguous stimuli in the middle of the 11 stimulus series. This reduced proportion of "row" responses to the ambiguous stimuli means that one would have to go further toward

the column end of the stimulus continuum before reaching the point at which 50% of the responses would be "rows" and 50% "columns." This would result in the shift in the indifference point that Bell and Bevan found.

We urge that fuller examination be given to the role played by the equal response principle in various experiments in which psychometric functions are obtained. Such functions, whether obtained by the method of constant stimuli, limits, or adjustment, may be contaminated by the artifactual effects of S's tendency to use responses with approximately equal frequency.

REFERENCES

BELL, R. A., & BEVAN, W. Influence of anchors upon the operation of certain Gestalt organizing principles. *Journal of Experimental Psychology*, 1968, 78, 670-678.
BLAKEMORE, C., & SUTTON, P. Size

adaptation: A new aftereffect. *Science*, 1969, 166, 245-247.
HELD, R. Adaptation to rearrangement and visual-spatial aftereffects. *Psychologische Beiträge*, 1962, 6, 439-450.
HELSON, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 1. New York: McGraw-Hill, 1959. Pp. 565-621.
HELSON, H. *Adaptation level theory*. New York: Harper & Row, 1964.
HELSON, H. Some problems in motivation from the point of view of the theory of adaptation level. In D. Levine (Ed.), *Nebraska symposium on motivation*. Lincoln: University of Nebraska Press, 1966. Pp. 137-182.
HOULIHAN, K., & SEKULER, R. Contour interaction in visual masking. *Journal of Experimental Psychology*, 1968, 77, 281-285.
LEVISON, M., & RESTLE, F. Invalid results from the method of constant stimuli. *Perception & Psychophysics*, 1968, 4, 121-122.
PANTLE, A., & SEKULER, R. Size-detecting mechanisms in human vision. *Science*, 1968, 162, 1146-1148.

PANTLE, A., & SEKULER, R. Contrast response of human visual mechanisms sensitive to orientation and direction of motion. *Vision Research*, 1969, 9, 397-406.
PARDUCCI, A. Category judgment: A range-frequency model. *Psychological Review*, 1965, 72, 407-418.
PARDUCCI, A., & HAUGEN, R. The frequency principle for comparative judgments. *Perception & Psychophysics*, 1967, 2, 81-82.
POULTON, E. C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, 1968, 69, 1-19.
RESTLE, F. Moon illusion explained on the basis of relative size. *Science*, 1970, 167, 1092-1096.
STEVENS, S. S. Adaptation-level vs. the relativity of judgment. *American Journal of Psychology*, 1958, 71, 633-646.
TUNE, G. S. Response preference: A review of some relevant literature. *Psychological Bulletin*, 1964, 61, 286-302.

(Accepted for publication July 6, 1970.)