AN EXPERIMENTAL COMPARISON OF THE METHOD OF LIMITS AND THE DOUBLE STAIRCASE-METHOD

By JACOB NACHMIAS and ROBERT M. STEINMAN,
University of Pennsylvania

It is a well-known psychophysical fact that the values of thresholds and *PSE*s depend in part upon the method of measurement.[1] In the method of constant stimuli, the distribution of values of the stimulus-variable presented to *O* affects at least measures of central tendency. In the traditional method of limits, the end-point of a run is undoubtedly biased by its starting point. This bias may be particularly serious when only ascending or only descending runs are employed, and when the variability of the end-points is a function of the parameter under investigation. It may even be present, however, when the two kinds of runs are interspersed, because the amount of the bias probably depends on the distance separating the starting point and the 'true' threshold or *PSE*. Thus, the method of limits permits *E* to influence his results, even if only inadvertently.

Cornsweet has recently described a variant of the staircase-method (itself a variant of the method of limits) which appears to be largely free of starting point bias.[2] In the usual staircase-method, or method of "up and down,"[3] *E* changes the value of the independent stimulus-variable by the same amount on every trial, but in a direction completely determined by *O*'s response on the just preceding trial: The direction of stimulus-change is reversed when *O*'s response changes, but not otherwise. In the new version, trials from two such staircase-series with different starting points are randomly mixed. This paper presents data from an experiment on the detection of a line-increment, designed to compare Cornsweet's double stair-

* Received for publication August 28, 1963. This investigation was supported by research grant B-3682(C1) from the National Institute of Neurological Diseases and Blindness, Public Health Service.

[1] R. S. Woodworth and Harold Schlosberg, *Experimental Psychology*, 1954, 225-233; W. S. Verplanck and J. W. Cotton, The dependence of frequencies of seeing on procedural variables: I. Direction and length of series of intensity-ordered stimuli, *J. gen. Psychol.*, 53, 1955, 37-47.

[2] T. N. Cornsweet, The staircase-method in psychophysics, this JOURNAL, 75, 1962, 485-491.

[3] W. J. Dixon and F. J. Massey, *Introduction to Statistical Analysis*, 1957, 279-286.

case-method and the traditional method of limits with regard to bias introduced by the starting point.

Experimental conditions. *O* was dark-adapted for at least 15 min., after which he looked at a small red fixation-point, while steadying his head with the help of a dental impression-plate. The stimulus-object consisted of a bright vertical line 2′ of arc wide superimposed on a disk of white light 1° in diameter and located 10′ of arc above the fixation-point. The disk and line appeared together for 52.4 m.sec. every 10 sec.; a warning buzzer preceded each trial by 2 sec. The luminance of the disk was fixed at 0.42 log m.L. while that of the line increment was varied in 1.1-db. steps in accordance with the particular psychophysical method employed. The details of the optical system, shutter-system, and method of calibration have all been reported previously.[4]

Observers. Both *O*s were men with at least 20:20 vision, uncorrected. One of them (*RS*) was the junior author, who knew well both the psychophysical methods employed and the purpose of this investigation. He was told, however, neither how many different starting points were to be tested nor the starting point of any particular series. The other *O* (*VV*) was told nothing about the methods or aims of the experiment, nor did he discover what they were during the course of the experiment.

Procedures. Each experimental session consisted either of 8 or of 4 double-staircase series and a like number of pairs of limits. The methods were alternated within each session, beginning with double staircases on half of the sessions. As shown in Table I, four different pairs of starting luminances of the line-increment were employed with both methods: two were high bias (*H* and *H′*) and two were low bias (*L* and *L′*). All four pairs were tested equally often in each session, but in a different, scrambled order. Furthermore, the lower of each pair of starting luminances was shown first randomly half the time.

The reference-level around which the starting points were centered was determined separately for each *O* on the basis of his performance in a few preliminary sessions. In fact, *RS*'s reference-level turned out to be 2.4 db. lower than *VV*'s (−0.09 log m.L. vs 0.15 log m.L.). After the reference-level was determined, both *O*s participated in several experimental sessions. The data reported are based on the last seven presentations of the four bias-conditions with each of the methods.

Every run by the method of limits terminated as soon as *O* changed his response. Every double staircase lasted for 32 trials. These trials were not random mixtures of two staircase-series, as Cornsweet recommends.[5] Rather, random permutations of the numbers 1 to 16 were used in such a way as to insure that the two staircases were represented by equal numbers of trials in the first and second halves of each double staircase-series.[6] This procedure made it possible to compare *O*'s performance on the first and second 16 trials without having to take account of the relative number of trials coming from the two staircases.

Results and discussion. The end-point of a run by the method of limits

[4] Jacob Nachmias and R. M. Steinman, Study of absolute visual detection by the rating scale method, *J. opt. Soc. Amer.*, 53, 1963, 1206-1213.

[5] Cornsweet, *op. cit.*, 489-491.

[6] W. G. Cochran and G. M. Cox, *Experimental Designs*, 2nd ed., 1957, 583-595.

was taken to be the decibel-mean of the luminance of the trial on which *O*'s response changed and the luminance on the just preceding trial. Similar means were obtained in double staircase-series, except that the just preceding trial from the appropriate staircase was considered, even if several trials intervened from the other staircase. For every occasion on which each of the four pairs of starting points was used, the means of the following end-points were computed: *double staircases*—(a) the first from each staircase; (b) the last from each staircase; (c) all of those occurring between Trials 1 and 16; (d) all of those occurring between Trials 16 and

TABLE I

STARTING LUMINANCES AND GRAND MEANS OF END-POINT LUMINANCES, EXPRESSED IN DECIBELS RE REFERENCE-LEVEL

Methods	*O*	Starting points			
		high bias		low bias	
		H	*H'*	*L*	*L'*
		5.10–1.67	4.00–0.55	1.75–4.88	0.57–3.80
Limits	*RS*	2.10	2.78	1.29	1.38
	VV	2.42	2.25	1.56	2.05
Staircase:					
First two	*RS*	2.45	2.19	0.00	1.01
	VV	2.40	2.63	2.21	0.84
Last two	*RS*	2.28	1.26	2.10	2.03
	VV	2.35	2.34	2.28	2.80
Trials 1–16	*RS*	2.62	2.04	1.01	1.10
	VV	2.52	2.65	2.19	1.24
Trials 17–32	*RS*	1.85	2.04	1.86	2.02
	VV	2.79	2.43	2.33	2.68

32; *limits*—the two end-points. Table I contains the grand means of these five measures.

The starting point clearly affects the average end-point in the method of limits, in agreement with the previous findings of Verplanck and Cotton.[7] An analysis of variance of these data revealed a significant effect of starting point ($F = 7.31$, $df = 1/52$, $P < 0.01$). Both *O*s were affected in much the same way: the *O* by starting point interaction was not statistically significant. The error-variance was 1.49 db.[2] In the double staircase-method, the effect is, if anything, more pronounced, if one considers only the end-points within the first 16 trials. Separate analyses of

[7] Verplanck and Cotton, *op. cit.*, 37-42.

variance on the end-points from the first and last 16 trials were performed. For these analyses, the two pairs of high bias starting points (H and H') were considered together, as were L and L'. The effect of the starting points on the end-points was statistically significant only in the first 16 trials ($F = 13.75$, $df = 1/52$, $P < 0.001$). The O by starting point interaction was not significant in either analysis. The error-variance in the analyses of the first 16 trials was 1.21 db.^{2} and in the last 16 trials, 0.97 db.^{2} The magnitude of the bias is not very large (about 1.0 db.), but it constitutes 25% of the decibel-difference between the mean starting points of the H-H' series and the L-L' series. Larger differences between mean starting points possibly would have produced larger effects.

On the other hand, two other measures from the double staircase-method appear to be unaffected by the locations of the starting point; namely, the last end-points and the mean end-point in the last 16 trials. These same measures, however, may not escape contamination if mean differences between starting points larger than 3.3 db. are tested. Since this psychophysical method will generally be used to find the value of an unknown threshold or *PSE*, it may be necessary to proceed beyond 32 trials before end-points reach a steady level unaffected by the starting points.