Title
The staircrase-method in psychophysics.

Journal
The American Journal of Psychology, 75(3)

ISSN
0002-9556

Author
Cornsweet, Tom N

Publication Date
1962-09-01

DOI
10.2307/1419876

NOTES AND DISCUSSIONS

THE STAIRCASE-METHOD IN PSYCHOPHYSICS

A psychophysical method variously referred to as the "method of up and downs,"[1] the Békésy audiometric method,[2] or the staircase-method, has come into extensive use in the last few years. The method has several advantages over other more commonly used techniques but it also has some disadvantages. This paper will illustrate the use of the method, will discuss its relative merits and demerits, and will describe a modification which overcomes certain of the disadvantages of the method.

The staircase-method is best described by illustrating its use with a specific problem. Suppose the problem is to determine *S*'s absolute, intensive threshold for the sound of a click. The first stimulus that *E* delivers is a click of some arbitrary intensity. *S* responds either that he did or did not hear it. If *S* says 'yes' (he did hear it), the next stimulus is made less intense, and if *S* says 'no,' the second stimulus is made more intense. If *S* responds 'yes' to the second stimulus, the third is made less intense, and if he says 'no,' it is made more intense. This procedure is simply continued until some predetermined criterion or 'number of trials' is reached. The results of a series of 30 trials are shown in Fig. 1. The results may be recorded directly on graph-paper; doing so helps *E* keep the procedure straight.

There are a number of ways of determining the intensive value that represents the threshold. The simplest is to compute the mean of the values of a given number of stimuli delivered after the series has reached its final level. This requires an arbitrary decision about when the final level has been reached. The technique, which avoids this difficulty and yields a 50% value, is simply to determine the stimulus above which 50% of the responses are 'yes,'—*i.e.* in Fig. 1 between 61 and 62 db.

Statistical treatment of the results has been discussed by Dixon and Massey, who describe the techniques for determining the means, standard deviations, standard errors, etc., for this type of data.[3] The treatments assume, however, that the response to each stimulus is independent of the preceding stimuli and preceding responses. This assumption holds for the examples analyzed, but there is evidence that the assumption does not always hold for human *Ss* in psychophysical experiments.[4] The development of techniques that take the existing inter-actions into account has not as yet been achieved.

---

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

[2] Georg von Békésy, A new audiometer, *Acta Oto-laryngol.,* 35, 1947, 411-422.

[3] Dixon and Massey, *op. cit.,* 286.

[4] W. S. Verplanck, G. H. Collier, and J. W. Cotton, Nonindependence of successive responses in measurement of the visual threshold, *J. exp. Psychol.,* 42, 1952, 273-282; Verplanck and 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; V. L. Senders, Further analysis of response sequences in the setting of a psychophysical experiment, this JOURNAL, 66, 1953, 215-229; R. S. Woodworth and Harold Schlosberg, *Experimental Psychology,* 1954, 225-233.

An important characteristic of the staircase-method is that *E* imposes four predetermined conditions upon every series. He decides in advance (1) where to start the series; (2) how large the steps are; (3) when the series should be stopped; and (4) when the series should be modified.

(1) Where to start. As in most psychophysical methods, several preliminary series are necessary to establish the range in which the threshold falls. In general the method is most efficient if the first stimulus is so near the threshold-level that many trials are not required in approaching the threshold.

(2) Step-size. Stimulus-intensities should be chosen along some scale that yields approximately equal sensory intervals. Logarithmic steps suffice for most applications of the method. The size of the steps should be so selected that no more than two, three, or four like responses are made before a reversal of the responses occurs.

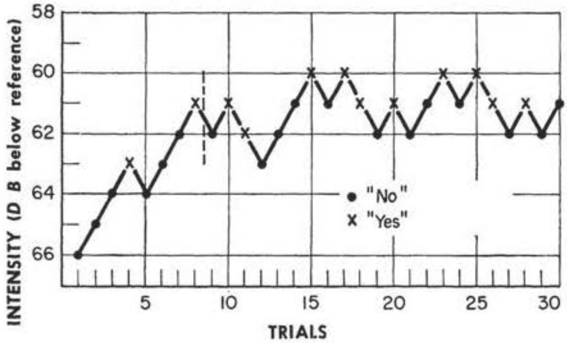

FIG. 1. DATA FROM THE DETERMINATION OF A TYPICAL AUDITORY THRESHOLD BY THE STAIRCASE-METHOD

For instance, if the step is very large, *S* will simply alternate 'yes' and 'no' as the intensity of the stimulus jumps from above to below his threshold. In this case, the value of the threshold would be insensitive to changes in the experimental variables. If, on the other hand, the step-size is very small, *S* will give long series of 'yes' or 'no.' When this happens, the method is inefficient because the succeeding stimulus-responses do not give as much information as they would if the step-size were larger. That is, if *S* has already responded with a 'no' (or 'yes') to a given intensity, the fact that he says 'no' (or 'yes') to an intensity only very slightly greater (or smaller) could have been predicted with a large degree of certainty. Therefore, when he actually does say 'no,' (or 'yes') not much new information is received. In general, this psychophysical method, as well as all others, becomes maximally efficient when the stimulus-steps are the size of the differential threshold.[5] Should this information be available, preliminary experimenting with step-size is unnecessary.

(3) When to stop. In general, the results with the staircase-method are like those shown in Fig. 1. The values of the stimuli presented change relatively rapidly un-

[5] Dixon and Massey, *op. cit.*, 289.

til they reach an asymptotic level or plateau, and then they hover around this level as long as the conditions remain unchanged. Obviously, the longer the series, the more reliable will be the computed value of the threshold. Thus, the decision about when to end the series is a compromise between a large number of trials for reliability and a small number for economy in time. The particular compromise depends upon the particular requirements of the experiment. The simplest method of determining when to end the series is to decide to conduct a predetermined number of trials. There are more efficient procedures, however, which take account of the fact that the values of the first several trials rest to a large degree upon where *E* began the series. These trials should not, therefore, be included in the final computation of the threshold. A better method is to decide, in advance, to end the series at a predetermined number of stimulus-responses after the plateau

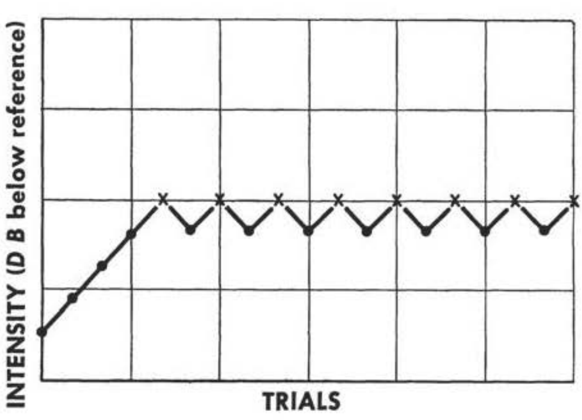

FIG. 2. DATA FROM A THRESHOLD-DETERMINATION UNDER IDEALIZED CONDITIONS

has been reached, and to include only these latter responses in the computation of the threshold-value. The problem, then, is to decide when the final level has been reached.

If there were no variability in the apparatus or in *S*, and if every response was completely independent of the previous ones, the data would look like those shown in Fig. 2 and *E* could simply say that the final level was reached at the intensity where the first reversal of response occurred. Data from human *S*s, however, show evidence both of variability and of interdependence of responses and a plot of the results usually obtained is like that shown in Fig. 1. To allow for variability in the apparatus and also in the *S*, we may decide that the final level has been reached after some arbitrary number of reversals greater than one has been reached. The particular number of reversals chosen depends upon the amount of variability and interdependency that occur under the particular conditions of the experiment, and this must be discovered in preliminary trials. For example, if the data in Fig. 1 were the result of preliminary trials, we should decide that the experimental run will consist of 10 stimuli delivered after the third reversal—after the dashed vertical line in Fig. 1.

(4) Modification of step-sizes. Under some conditions, the size of the steps in

staircase-method should be varied during the course of the experiment. For example, if the method were used to determine the absolute, visual, intensive threshold, the steps in the preliminary stage of dark-adaptation should be large and in the final stage they should be much smaller. From the point of view of careful experimental design, it would be desirable that the first stimulus in each of the staircases be at the same intensity-level. As a result, in at least one of the series, and probably in both, the starting level would be far removed from the final level. This may be avoided by using large steps in getting to the final level and then switching to smaller steps when the final level has been reached. For example, it may be decided in advance that, for the visual-threshold, the steps will be 50% changes in intensity until the third reversal, and 10% changes thereafter. As a rule of thumb, when this sort of procedure is used, a series that contains a total of 20–25 trials will produce a fairly reliable threshold-value under a number of diverse conditions.

Advantages of the staircase-method. The staircase-method is extremely efficient. For a given reliability of a computed threshold-value, the staircase-method requires the presentation of many fewer stimuli than any other psychophysical method, because, once the first few stimuli are out of the way, all of the other stimuli are very near the threshold-level, each one contributing importantly to the final computed threshold-value.

Comparison with method of adjustment. A comparison of the staircase-method with the method of adjustment is difficult. The two methods are very similar, since, in the method of adjustment, *S* changes the stimuli in much the same way that *E* does in the staircase-method. To the extent that this is true, the two methods may be equally efficient. The staircase-method has, however, one advantage in that there is no ambiguity about what is actually happening. With the method of adjustment, the particular values of the stimuli that *S* has experienced are usually unknown, and he process by which *S* decides what value to settle for is certainly less clear than the corresponding method used by *E* in computing the threshold from a set of data obtained by the staircase-method.

In addition to its efficiency, the staircase-method is very convenient for tracing the temporal course of changes in a threshold when these changes are reasonably slow. For example, the method has been used to great advantage by Blough in plotting the temporal course of dark-adaptation in pigeons.[6]

There is a theoretical consideration which reveals a very important property of the staircase-method. An *S*'s response to a given stimulus in a series of stimuli depends not only upon the value of the given stimulus, but also to some degree upon the values of all of the stimuli that have already been presented, and upon the responses he has given to these stimuli. This fact results in phenomena variously called anchoring effects, series-effects, adaptation-level phenomena, perseveration, anticipation, etc.[7] As a consequence, the measured value of a threshold or a point

[6] D. S. Blough, Dark adaptation in the pigeon, *J. comp. physiol. Psychol.*, 49, 1956, 525-530; Method for tracing dark adaptation in the pigeon, *Science*, 121, 1955, 703-704.

[7] See Woodworth and Schlosberg, *op. cit.*, 225-233 for a general discussion of these effects.

of subjective equality (*PSE*) depends to some degree on the method of measurement employed.[8]

The dependence of these measures (threshold and *PSE*) upon the method used must be carefully considered in the design of an experiment. For instance, consider an experiment to determine the effect of Condition X upon a *PSE*. Two *PSE*s will be determined, one with and one without Condition X. It is obvious that the same method, and in fact the identical measuring conditions (*e.g.* the same set of stimuli for the method of constant stimuli) must be used in determining each *PSE* if a discovered difference is to be attributed to Condition X. But it is evidently not so commonly acknowledged that the actual magnitude of the difference between the two *PSEs* will be influenced by the measuring method even when the two conditions are made identical. It is usually found that serial effects result in the *PSE*s being pulled toward the middle of the set of stimuli that are presented.[9] Therefore, if the two sets of stimuli delivered under the two conditions are the same, each of the *PSEs* will be pulled toward the *same* value by the serial effects. The measured difference between *PSEs* must therefore become smaller as the interactions between responses become greater.

Disadvantages. When the staircase-method is used, *S* is very much aware of the way in which the stimuli are being ordered. Even if he is naïve to begin with, it does not take him long to become aware of the procedure. If the judgments are easy this awareness is not disturbing, but as soon as the judgments become difficult, as they are in almost all psychophysical research, it becomes distressing.

When *S* has reported 'no' four successive times, his attitude and criteria on the next trial are not at all what they were at the beginning of the series. This is, of course, a reflection of the inter-dependencies of the series. Another related consequence of the series inter-dependencies is the fact that *S* can, if he pleases, manipulate the results at will. If he wished to 'cheat,' he could at any point in a series of stimuli and continue to give what look like meaningful data. The fact that it is possible for *S* willfully to malinger means that the results may also be influenced by the biases of an 'honest' *S*. This factor becomes crucial when effects of small magnitude are being studied, particularly when the *S*s are acquainted with the purposes of the study. In many psychophysical studies, it is desirable to gather very large quantities of data on a few individuals, and, particularly when those individuals serve as *E* as well as *S*, the possible effects of *S*'s biases can be severe enough to offset the advantages of the staircase-method.

The double staircase-method. To reduce the interdependencies and to

---

[8] Some of the forced-choice methods may be exceptions to this rule. (H. R. Blackwell, Contrast thresholds of the human eye, *J. opt. Soc. Amer.*, 1946, 36, 624-643). Heinemann has developed an ingenious way to use the staircase-method, in combination with a forced-choice procedure for the determination of differential thresholds (E. G. Heinemann, The relation of apparent brightness to the threshold for differences in luminance, *J. exp. Psychol.*, 61, 1961, 389-399).

[9] Woodworth and Schlosberg, *op. cit.*, 225-233.

lessen the possibilities that they may bias the results, stimuli chosen by some procedure different from the procedure that determines the regular stimuli, may be inserted at various points in the series.

One way of doing this is to scatter stimuli throughout the series with randomly selected values. Such a procedure would, however, have two disadvantages: (1) it would lose efficiency; and (2) the values of the non-staircased stimuli would so influence the final level that their selection would reintroduce artifacts of anchoring. A better solution is to run two series of staircased stimuli concurrently. The data resulting from a simple alternation of two staircases are shown in Fig. 3. The

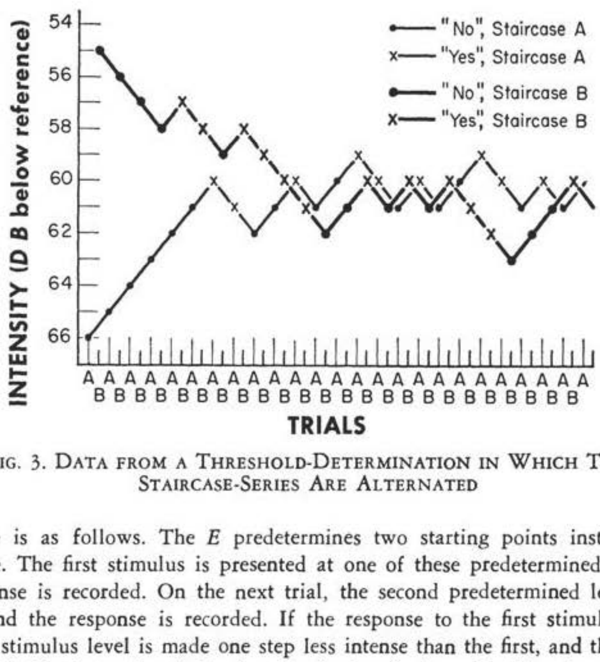

FIG. 3. DATA FROM A THRESHOLD-DETERMINATION IN WHICH TWO STAIRCASE-SERIES ARE ALTERNATED

procedure is as follows. The *E* predetermines two starting points instead of the usual one. The first stimulus is presented at one of these predetermined levels and the response is recorded. On the next trial, the second predetermined level is presented, and the response is recorded. If the response to the first stimulus is 'yes,' the third stimulus level is made one step less intense than the first, and the response is recorded. The intensity of the fourth stimulus is determined by the response to the second stimulus, and so on. In this way, two staircase-series are run concurrently, one on odd- and the other on even-numbered trials, each alternate stimulus depending upon the response to the previous stimulus in its own staircase.

Now go one step further. Let us call one staircase *A*, the other *B*. Instead of running two series, the *A* on odd and the *B* on even trials, the choice as to which staircase will be represented on any given trial is made randomly. Data collected from such a double randomly intermixed staircase are shown in Fig. 4. The order of occurrence of the staircases was chosen in advance from a table of random numbers. In general, if the two series are started apart as in this figure, they will come together and then cross and recross each other thereafter. The data to be used in the final computation of the threshold-value may be chosen in just the same way as they are for a single staircase. A double run may be considered either as two replications of the same condition, or the data may be combined and a single threshold-value computed.

When the random, double staircase-method is used, S feels none of the constraint that goes with the single staircase-method. The difference is perhaps best illustrated by the fact that it is no longer possible for S to close his eyes and give meaningful data. As long as the choice of staircase A or B is made randomly on each trial, there is no strategy of responding that S

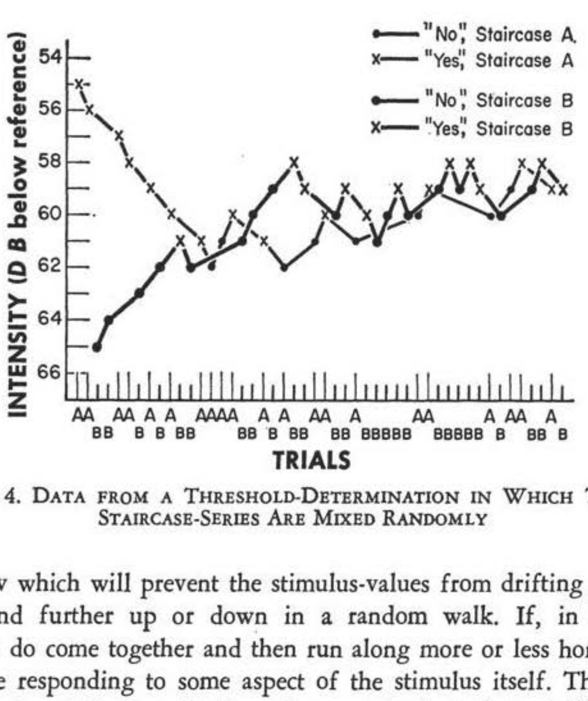

FIG. 4. DATA FROM A THRESHOLD-DETERMINATION IN WHICH TWO STAIRCASE-SERIES ARE MIXED RANDOMLY

can follow which will prevent the stimulus-values from drifting gradually further and further up or down in a random walk. If, in fact, the two series do come together and then run along more or less horizontally, S *must* be responding to some aspect of the stimulus itself. There is no other way in which the stimulus-values can be kept from drifting. This means that the possibility of the effects of series- interdependencies biasing the computed threshold-value is greatly reduced.

In view of all the arguments presented above, the double-staircase method provides a means of determining psychophysical functions that has the practical advantage of high efficiency while retaining the theorectical advantage that the final results are relatively little influenced by the biases that E must introduce into any experiment.

University of California, Berkeley TOM N. CORNSWEET