

A sequential test of performance (STOP) for all-or-none learning data¹

ALLAN J. NASH, Florida Atlantic University, Boca Raton, Fla. 33432

A procedure is developed using the binomial test for the sequential criterion testing of all-or-none learning data. The sequential test is accompanied by a format for gathering the data which generates a graphic profile of S's performance over trials. Comparison of an index of problem difficulty provided by the Sequential Test of Performance (STOP) with a more commonly used index, the number of correct responses in a fixed block of trials, produced a high linear relationship between the two measures with the STOP method showing a savings in trials run of over 40%.

METHOD

The criterion for learning described in this paper is particularly efficient in situations where a two-state, all-or-none learning process, similar to those described by Theios (1968) can be assumed. Specifically, the method assumes that (a) responses can be classified as either correct or incorrect on each of a series of trials, (b) on any trial prior to learning, there is a constant probability of a chance-produced correct response (in this paper this probability is, for convenience, set equal to 0.5), and (c) when learning occurs the probability of a correct response takes a single jump to an asymptotic value, equal to or less than unity.

The usual procedure for establishing that learning has taken place under an all-or-none model terminates testing after a sequence of, say, 10 or more consecutive correct responses. This method is satisfactory when the probability of a correct response in the learned state is close to unity, but as Theios (1968) points out, there are many situations, particularly in animal studies, where the asymptotic correct response probability is less than unity. In such cases the persistence of errors in the learned state requires a different criterion for terminating testing. The present procedure is appropriate in such instances.

THE SEQUENTIAL TEST OF PERFORMANCE (STOP)

The STOP method for specifying when a criterion is reached is equivalent to running a binomial test after each trial on the hypothesis (H_0) that the probability of a correct response = $\frac{1}{2}$. The critical numbers of binomial responses required to reject H_0 for $p = \frac{1}{2}$ and sample sizes up to 1000 may be obtained from a table of critical values prepared by MacKinnon (1964). These critical values are graphically summarized in Fig. 1.

The data sheet for STOP is laid out (after Bross, 1965, p. 138) as a coordinate system which shows when a criterion is reached. The starting point for any problem series has coordinates, (0,0), and is marked by an X on the data record in Fig. 1. If a correct response is made on the first trial, the point directly above the starting point is encircled. If an incorrect response is made on the first trial, the point directly to the right of the starting point is encircled. These rules apply to all subsequent trials: for every correct response, the point directly above the last one marked is encircled; for every incorrect response, the point directly to the right of the last one marked is encircled. In Fig. 1, the sample data record shows that the first five trials were classified as "correct, incorrect, incorrect, correct, incorrect . . ." Testing is continued until a criterion barrier is crossed. The .05 criterion barriers are shown in Fig. 1 by single-width lines. The .01 criterion barriers are given by double-width lines. When a decision barrier is crossed, an index of problem difficulty may

Sequential Test of Performance (STOP)

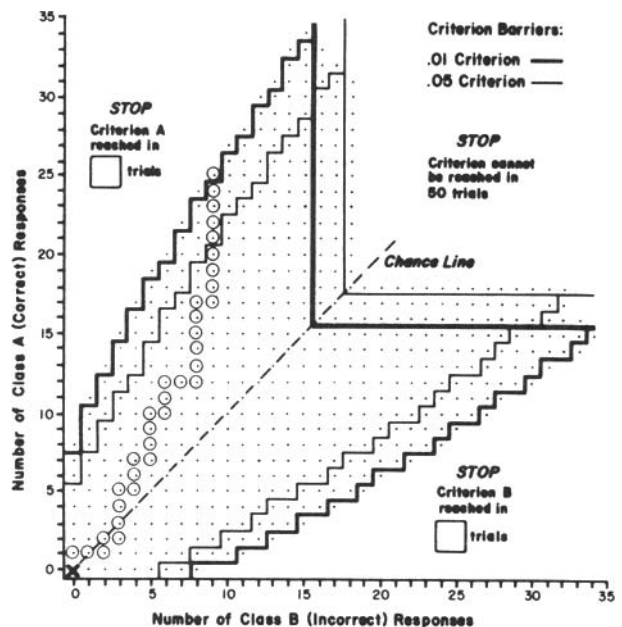


Fig. 1. STOP format for recording observations.

be defined by the number of trials required to reach that criterion. This index is given by the sum of the coordinates at that point. The barrier for incorrect responses will seldom be crossed in most learning experiments but its inclusion broadens the applicability of the STOP method to situations where either of the two response classes may potentially occur with sufficient frequency to warrant rejecting H_0 .

The right-angle barriers that cut pie-shaped segments out of the upper right portion of Fig. 1 are also stopping points. When one of these barriers is crossed it means that it is impossible to reach the corresponding criterion level in 50 trials. If this barrier is crossed and a 50-trial maximum has been set by E, testing should be stopped at that point since S cannot reach criterion in less than 50 trials no matter how he responds after crossing the barrier. If many Ss are likely to require more than 50 trials to reach criterion, a similar data-sheet format which extends up to a 100-trial maximum can be used as a continuation of the form given in Fig. 1.

The broken line extending at a 45-deg angle from the origin in Fig. 1 represents the expected value of chance performance at any point in the series of trials. The STOP method for recording the data provides an immediate graphical representation of S's trial-by-trial performance. It is particularly helpful in spotting sequential properties of the data.

COMPARISON OF STOP WITH ANOTHER INDEX OF PROBLEM DIFFICULTY

The two-choice discrimination learning data reported by Nash & Michels (1966) were reanalyzed over the first 50 trials using the STOP criterion. For each of 10 Ss, the mean number of trials to criterion (STOP) over eight different discrimination tasks was computed and correlated with the mean number of correct responses in the first 50 trials on the same tasks.

(Continued on page 79)

trend analysis also showed that the linear slopes of the sex groups were significantly different ($F = 5.88$, $df = 1/90$, $p < .025$), that of the females being steeper than that of the males (-1.26 vs $-.80$, respectively).

A mixed analysis of Trials 9 and 10 revealed that fixation recovery on Trial 10 was significantly different for the three stimulus groups ($F = 15.57$, $df = 1/90$, $p < .001$). The increment for the 2×2 group was greater than that for both the 12×12 and 24×24 groups ($p < .01$ in each case), but the latter two did not differ from each other. Wilcoxon tests indicated that the increment for each individual stimulus and sex group was significant ($p < .05$ for the 24×24 group; $p < .01$ for all other groups), thus demonstrating that the preceding decrements were due neither to general nor to sensory fatigue.

The reappearance of the previously repeated stimulus (Trial 13) again produced differential and transitively ordered decrements in TF. A two-factor analysis of the relative decrease from Trial 5 to Trial 13 [(5 - 13)/5] yielded significant main effects for Stimulus ($F = 9.79$, $df = 2/90$, $p < .001$) and for Sex ($F = 10.42$, $df = 1/90$, $p < .005$). Thus, the 2×2 group, as expected, again declined more steeply than both the 12×12 and 24×24 groups ($p < .05$ and $p < .01$, respectively) and the 12×12 group showed an intermediate decrease (though not quite significantly different from the 24×24 group). The decline of the girls here, also, was steeper than that of the boys.

Finally, analysis of Trials 13 and 14 indicated that recovery on Trial 14 was significantly different for the three stimulus groups ($F = 19.02$, $df = 2/90$, $p < .001$) and for the two sex groups ($F = 11.28$, $df = 1/90$, $p < .005$). The increment for the 2×2 group was, as before, greater than that for both the 12×12 and 24×24 groups ($p < .01$, in each instance) and that for the 12×12 group, greater than that for the 24×24 group ($p < .01$). The increment for the females was larger than that for the males. Wilcoxon tests again demonstrated that recovery was significant for each stimulus and each sex group ($p < .01$, in each case).

DISCUSSION

The results, taken together with those of Brennan, Ames, & Moore (1966), suggest that the more complex a checkerboard

pattern the more likely it is not only to recruit but also to maintain the attention of 14- to 15-week-old infants. At least this is the case for the range of density values presently employed. The fact that the habituation data were transitively ordered indicates, moreover, that a single physical dimension (correlated with checker density) was the effective controlling variable (cf. Hershenson, 1967). In addition, the marked differences in response to the repeated and novel stimuli subsequent to the repetition period attest to the discriminative competence of the infants.

Finally, the sex differences in magnitude of habituation merit comment. It is not uncommon to find infant girls described as showing greater differentiation between stimuli or as habituating more rapidly than boys. On this basis, they are often said to be "maturationally ahead" (cf. Kagan & Lewis, 1965; Kagan, Henker, Hen-Tov, Levine, & Lewis, 1966). It would seem premature, however, to interpret the steeper decline of the girls in this study as evidence of precocious perceptual development, since none of the many other possible determinants of these differences (e.g., sampling artifacts, differences in overall adaptation to the test situation, etc.) have been ruled out.

REFERENCES

- BRENNAN, W. M., AMES, E. W., & MOORE, R. W. Age differences in infants' attention to patterns of different complexities. *Science*, 1966, 151, 354-356.
- CARON, R. F., & CARON, A. The effects of repeated exposure and stimulus complexity on visual fixation in infants. *Psychonomic Science*, 1968, 10, 207-208.
- HERSHENSON, M. Development of the perception of form. *Psychological Bulletin*, 1967, 67, 326-336.
- KAGAN, J., & LEWIS, M. Studies of attention in the human infant. *Merrill-Palmer Quarterly*, 1965, 11, 95-127.
- KAGAN, J., HENKER, B., HEN-TOV, A., LEVINE, J., & LEWIS, M. Infants' differential reactions to familiar and distorted faces. *Child Development*, 1966, 37, 519-532.

NOTES

1. Supported in part by USPHS Grant HD 01605-3 from NICHD. The assistance of Roberta Caldwell and Elizabeth Bond is gratefully acknowledged.

2. Preliminary analyses indicated that the performance of the new S_s in the 2×2 and 12×12 groups was no different from that of infants tested earlier.

(Continued from page 77)

The high linear relationship between the two measures ($r = -.949$) suggests that the use of the STOP method as the index of problem difficulty would lead to conclusions that do not differ markedly from those reached using the number of correct responses in 50 trials. The estimated mean savings in trials run per S using the STOP criterion was 20.8, a savings of more than 40% of the trials run under the other method.

SOME FINAL CONSIDERATIONS

One practical limitation on the procedure is that tables of critical values for the binomial test are not readily available for probabilities other than $\frac{1}{2}$, particularly for trial lengths over 25. A similar practical restriction applies to extending the present procedure beyond a 100-trial maximum. The suggested format for the data sheets, with their broadening pattern of points and barriers, would make it difficult to construct acceptable data sheets for trial lengths much beyond 100.

Finally, it should be noted that the terms, ".05 or .01 levels of significance," should be avoided when interpreting the criterion level of the STOP method. The binomial test is ordinarily applied to a single complete sequence of observations rather than to progressively larger subsequences. It is true in the present usage of the binomial test that when a .05 (or .01) criterion barrier is crossed, the probability of this

level of performance occurring by chance in exactly that many trials is less than .05 (or .01), but the STOP procedure does not specify the exact trial on which a barrier will be crossed. Consequently, the probability that S , by chance, would cross a criterion barrier at *any* one of the possible crossing points is greater than the level of significance used for establishing the individual criterion points. The practical effect of using the binomial test in the present manner is to control Type I errors by requiring S to reach a progressively more stringent criterion as the number of trials run increases.

REFERENCES

- BROSS, I. D. J. *Design for decision*. New York: The Free Press, 1965.
- MACKINNON, W. J. Table for both the sign test and distribution-free confidence intervals of the median for sample sizes to 1000. *Journal of the American Statistical Association*, 1964, 59, 935-956.
- NASH, A., & MICHELS, K. M. Squirrel monkeys and discrimination learning: Figural interactions, redundancies, and random shapes. *Journal of Experimental Psychology*, 1966, 72, 132-137.
- THEIOS, J. Finite integer models for learning in individual subjects. *Psychological Review*, 1968, 75, 292-307.

NOTE

1. This research was supported in part by Research Grant No. HD-1454 from the National Institute of Child Health and Human Development.