

Concept identification with reinforcement of hypotheses and misinformative stimulus labeling¹

JAMES R. ERICKSON and KAREN K. BLOCK, *Human Performance Center, Ohio State University, Columbus, Ohio 43210*

College students solved simple concept-identification problems in which feedback was contingent on their current hypothesis (H). No dimension was consistently labeled until S's pattern of responses over a block of stimuli indicated that he was using the solution H. The results indicated that some Ss were pure H testers, choosing new Hs on the basis of local consistency; these Ss had little difficulty with their problems. The remaining Ss evidently tried to integrate the unreliable feedback information over more than one trial and found their problems quite difficult or impossible to solve.

Recent research has shown that when Ss are carefully instructed as to the nature of the solution of simple one-dimensional concept-identification problems, and know what the possible solutions are, their responses while solving such problems are almost completely organized by hypotheses (Hs) about the solution. Furthermore, Ss seem to make fairly efficient use of at least the most recent stimulus-feedback pairing when choosing new Hs to test (Erickson, 1968; Levine, 1966). In these studies, Hs were inferred from S's response patterns over blocks of nonfeedback trials. At the end of each block one feedback trial was given, informing S as to the classification of the stimulus appearing on that trial.

In the present study Ss were instructed that this same procedure would be used—that he could compare his guess as to how the stimulus was classified with the way it was really classified and by using this information, solve the problem. However, this was not the case. If S's pattern of responses on the nonfeedback trials indicated that he was not using the solution H, the feedback always indicated that the stimulus was labeled the opposite of his response, indicating an error, but providing no reliable information as to the actual classification of the stimulus.

According to cue-conditioning (Bourne & Restle, 1959) models of concept identification, this procedure should make the problem quite difficult, if not insoluble, since none of the cues is consistently reinforced. The same is true for H-testing models which assume that S eliminates Hs from consideration by a consistency check procedure (eliminating dimensions which are inconsistently reinforced over two or more feedback trials). Such models have been proposed by Trabasso & Bower (1966) among others. On the other hand, pure sampling-with-replacement models, such as Bower & Trabasso (1964) and Restle (1962) have proposed, or models assuming memory for Hs previously tried-but-rejected, but not for stimulus classifications, would predict no particular difficulty with the above procedure, since only incorrect Hs would lead to error trials, and the correct H would be consistently reinforced.

One purpose of the present study was to examine behavior in a situation in which anything other than "pure" H sampling would cause difficulty. A second purpose was to examine resampling of Hs tried-but-rejected in this situation. Each S solved two problems, a control (C) problem in which the solution was chosen randomly and an experimental (E) problem where S's first H was called wrong and made the basis for solution from then on, a procedure used previously by Erickson (1968).

SUBJECTS

The Ss were 48 females from the introductory psychology S pool. An additional 20 Ss were run, but did not solve one of their problems.

PROCEDURE

The materials, procedures, counterbalancing controls, etc., were the same as those used in the Erickson (1968) study. Details may be found therein.

The Ss were given detailed instructions about the nature of the solutions of the problems, stressing that each stimulus would be labeled "A" or "B" according to the value of one of the dimensions. After solving two practice problems, each of which involved two binary dimensions, they were given instructions concerning the main problems, informing them that they would not receive feedback on every trial and that on nonfeedback trials they were to continue responding as if they had been correct. Before they solved each problem, Ss were shown examples of the stimuli to be used, and the various dimensions were pointed out.

The two main problems each contained four binary dimensions. For the first problem, each S's solution was randomly chosen, with the restriction that each of the eight possible solutions was used an equal number of times. For the second problem, S's first H was called wrong (by giving "error" feedback) and was then made the solution. Feedback was given according to the H inferred from S's response patterns on nonfeedback trials; only when the solution H was used was S's labeling response called correct (i.e., E's indication of the classification of a stimulus agreed with S's). The criterion for solution of the two main problems was four successive four-trial blocks on which S used the solution H.

RESULTS

The H-reinforcement procedure made the problem quite difficult. Compared with the Erickson (1968) study, which was similar except that reliable classification feedback was used, in the present study five times as many Ss failed to solve one of the problems (29% vs 6%), and the Ss who did solve both problems took roughly twice as many Hs to do so (9.9 vs 5.2). Preliminary analyses revealed that many of the Ss who solved their problems used several "impossible" Hs—response

Table 1
Summary Statistics for Various Groups of Subjects

Statistic	Problem	Hypothesis Reinforcement		Standard Reinforcement
		Hypothesis Testing Ss (N = 24)	Other Ss (N = 24)	Data from Erickson (1968)
Number of hypotheses to solution	C	5.71	12.67	4.10
	E	8.08	13.25	6.29
p(win-shift)	C	—	—	.02
	E	—	—	.02
p(lose-stay)	C	.01	.04	.02
	E	.02	.01	.02
p(local consistency)	C	.88	.59	.79
	E	.90	.75	.78
p(impossible hypothesis)	C	.05	.22	.02
	E	.03	.14	.02

patterns on nonfeedback trials which were not consistent with any of the eight possible solutions given by the instructions. The main results are given below separately for the Ss who had at most one such impossible H while solving each problem (called H-testing Ss henceforth) and the Ss who had more than one impossible H on at least one problem.

The basic summary data are shown in Table 1, along with comparable data from the Erickson (1968) study. The average number of Hs to solution (equivalent to errors or trials to criterion in terms of the basic results) for the H-testing Ss was comparable to that found in the earlier study. For these Ss the difference between the E and C problems was statistically significant ($F = 6.23$, $df = 1/23$; $p < .025$). The other Ss found the problems much more difficult, and the difference between E and C problems was not significant ($F < 1.00$). The mean difference of 2.3 Hs between E and C problems for H-testing Ss is similar to that found in the earlier study; several H-sampling models which could account for this difference are discussed therein.

The probability of keeping the same H following an error feedback is roughly the same in all cases, approximately .02. The probability of choosing another H following a "correct" feedback, .02 in the earlier study, is not meaningful in the present study, since Ss would be correct only if they used the solution H, and this was quite infrequent during the precriterion period. For only 25 out of the 952 precriterion Hs (over both problems) was the correct H used, reinforced positively, and then rejected by S; this is almost the same as the 2% "error rate" found with the other measures.

The probability of the H-testing Ss using an impossible H was of course low, since this was the basis of their classification. For the remaining Ss, this probability was quite high, about 10 times as high as was found with standard feedback. The probability dropped somewhat for the E problem, perhaps because Ss found they could solve the C problem using one of the Hs given in the instructions.

The probability of local consistency (choosing a new H which is consistent with the previous feedback) is an important statistic since several models have been proposed in which local consistency is assumed (Erickson, 1968; Gregg & Simon, 1967; Trabasso & Bower, 1966). There was a striking difference in local consistency between H-testing Ss and the other Ss, especially on the C problem. The Ss who tried impossible Hs had a local consistency probability near chance on the C problem; this increased to .75 on their second problem. The H-testing Ss were much more likely to choose locally consistent Hs. Their probability is equal to or higher than that found with standard feedback conditions. When calculated for the 13 Ss who had no impossible Hs, $p(\text{local consistency})$ was found to be even higher, about .95 for both problems.

It was thought that non-H-testing Ss might have had more dimensions inconsistently labeled than H-testing Ss, especially on early feedback trials. This could conceivably have caused the observed differences in $p(\text{local consistency})$. For the first nine pairs of feedback trials (1&2, 2&3, . . . 9&10), the mean number of dimensions classified differently on the two trials was calculated. There were no differences between the two types of Ss. For example, on the C problem H-testing and

non-H-testing Ss had an average of 2.23 and 2.26 dimensions inconsistently reinforced per feedback trial, respectively.

DISCUSSION

There were at least two strategies used by Ss in solving their problems. The strategy used by Ss who were most successful is close to that suggested by several current H-testing models of concept identification. These Ss used almost exclusively Hs which had been given as possible solutions, and chose new Hs to test on the basis of local consistency, evidently not remembering or ignoring inconsistent information from feedback trials other than the last one. These Ss evidently did remember Hs recently tried and rejected, and did not resample them immediately; this is reflected in the significant difference between C and E problems.

The remaining Ss evidently tried to integrate classification information over more than one feedback trial while solving their problems. Since the feedback labels actually given were associated with the true classifications only on a chance basis, such a strategy would quickly lead to inconsistencies, and undoubtedly caused confusion. Almost one-third of the Ss tested could not solve their problems; yet a simple concept-identification problem with four binary dimensions is not difficult in normal circumstances. The remaining Ss who did solve their problems had many response patterns not consistent with any of the possible solutions given in the instructions. It seems likely they tried Hs based on more than one dimension or based on stimulus sequences as they found that no one dimension was categorized consistently according to the feedback labels given. Once the C problem was solved, and the solution was one of those given in the instructions, the number of impossible Hs decreased and the probability of local consistency increased on the E problem.

It is apparent that none of the current theories of concept identification is universally applicable, but that each theory's usefulness will depend, for reasons not yet clearly understood, on situation-specific variables, including, it seems, individual-difference variables. Future theoretical and empirical work must be aimed at specifying the conditions under which various models will be applicable.

REFERENCES

- BOURNE, L. E., JR., & RESTLE, F. Mathematical theory of concept identification. *Psychological Review*, 1959, 66, 278-296.
- BOWER, G., & TRABASSO, T. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford, Calif.: Stanford University Press, 1964. Pp. 32-96.
- ERICKSON, J. R. Hypothesis sampling in concept identification. *Journal of Experimental Psychology*, 1968, 76, 12-18.
- GREGG, L. W., & SIMON, H. A. Process models and stochastic theories of simple concept formation. *Journal of Mathematical Psychology*, 1967, 4, 246-276.
- LEVINE, M. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, 1966, 71, 331-338.
- RESTLE, F. The selection of strategies in cue learning. *Psychological Review*, 1962, 69, 329-343.
- TRABASSO, T., & BOWER, G. H. Presolution dimensional shifts in concept identification: A test of the cue-sampling axiom. *Journal of Mathematical Psychology*, 1966, 3, 163-173.

NOTE

1. This research was carried out at the Human Performance Center and was supported in part by National Institutes of Health Grant No. MH-12113-02.