Some characteristics of hypothesis sampling in concept identification*

JAMES R. ERICKSON, KAREN K. BLOCK,[†] and MICHAEL J. RULON
Human Performance Center, Ohio State University, Columbus, Ohio 43210

Two studies were conducted in which Ss solved a concept-identification problem whose solution was a hypothesis (H) recently tried, but which had been incorrect when it was tried. In both experiments, the difficulty of the final problem was linearly related to the recency with which the solution H had previously been sampled. When the solution was an H that S had not as yet tried, problem difficulty was approximately the same as when the solution was the H three back from the start of the final problem. Analysis of the H-sampling data revealed that Ss made only partial use of information from previous stimuli when choosing a new H to test. A tentative model of H sampling is presented which takes account of these and other data on use of memory while solving concept-identification problems.

Most recent theories of concept identification have assumed that Ss solve concept-identification problems by sampling from a pool of hypotheses or possible solutions to the problem. Several alternative sampling schemes have been proposed (e.g., Bower & Trabasso, 1964; Erickson, 1968a, Gregg & Simon, 1967; Levine, 1966, 1969; Restle, 1962; Trabasso & Bower, 1968). This article presents data from two studies designed to provide information relevant to the sampling question. Both studies used the technique developed by Levine (1966) for inferring S's hypothesis (H) from his response pattern during a block of nonfeedback trials. Both studies involved tasks in which the solution was an H that S had recently tried but rejected because it led to an error. In Experiment 1, Ss first solved a concept-identification problem whose solution had been randomly chosen by E, then were shifted without warning to a second problem whose solution was an H they had tried recently or one they had not yet tested. Experiment 2 was similar, except that there was no preshift problem; instead, the first three Hs that S tried were called wrong, after which one of these three, or a new H, became the solution. Analysis of shift-problem difficulty as a function of how recently the "shift H" had been tried provides information concerning memory for Hs, and analysis of H sequences provides information on use of prior stimulus information.

EXPERIMENT 1

Method

The Ss were 173 female students from

*This research was carried out at the Human Performance Center and was supported in part by National Institutes of Health Grant No. MN-12113. A paper based on Experiment 1 was read at the Midwestern Psychological Association meetings, 1967.

[†]Now at the Learning Research and Development Center, Pittsburgh, Pa. 15213.

the psychology S pool. Of these Ss, 16 did not solve the first problem within the 1-h time period, and 24 did not solve the shift problem and were not included in the main analysis.

The stimuli were square figures varying along four binary dimensions: single or double border, solid or broken border, a line inside the square that was horizontal or vertical, and location of the inside line in the upper or lower half of the square. The stimulus materials were rear-projected onto a screen at the front of S's booth. The Ss sat at a table with a small response box containing two buttons and two feedback lights, each labeled "A" and "B."

The Ss were assigned randomly to four main groups, defined in terms of the shift-problem solution [Lag 1, Lag 2, Lag 3 H or new H (NH)]. A Lag 1 H is the H tried immediately before selecting the solution H of the first problem, the Lag 2 H is the one tried two back, etc., while an NH is one not previously sampled. The Ss who solved the first problem before trying three Hs and Ss for whom two of the Lag 1, 2, or 3 Hs were identical were assigned randomly to one of two other groups. The solution of the second problem for these two groups was an H that was either an intradimensional shift (IDS) or a randomly selected extradimensional shift (EDS) of the first problem solution. There were 16 Ss who had at least three different Hs immediately before solving their first problem in each lag group; separate analyses were performed for these Ss and for all Ss run. Final Ns for the various groups were: Lag 1, 24; Lag 2, 21; Lag 3, 24; NH, 16; IDS, 24; EDS, 24.

The Ss were given detailed instructions about the nature of the solution of the problems and solved two practice problems involving two binary dimensions to ensure that the procedures and the nature of the solution were understood. Before solving

each problem, S was shown examples of the stimuli to be used while the various dimensions were pointed out. For the final problem, in which the hypothesis-monitoring procedure was used, Ss were told that they would not receive feedback on every trial, and that when no feedback was given they were to continue responding as if they were correct. Feedback was given every third trial.

With four binary dimensions, 16 stimuli are possible. Five sets of 15 of these stimuli were divided into five three-trial blocks each, such that each block permitted the inference of a hypothesis (Erickson, 1968a; Levine, 1966). A different stimulus was omitted from each of the 15-slide sets. Orderings of the five sets were counterbalanced across Ss within conditions as much as possible.

The solution to the first problem was selected randomly for each S, with the restriction that each of the eight possible solutions was used approximately an equal number of times within each condition. After the first problem was solved to a criterion of two successive three-trial blocks in which the solution hypothesis was used, S was shifted without warning to her shift problem. The shift-problem criterion was six successive three-trial blocks in which the correct hypothesis was used.

On trials when feedback was given, it was given 2 sec after the response and remained on for 2 sec, after which both the feedback light and slide went off. Ten seconds later the next slide came on. When no feedback was given, the slide went off 4 sec after the response, and the intertrial interval was 10 sec.

Results

Figure 1 shows the number of three-trial blocks to solution for the various groups on the shift problem. On the basis of the assumption that Ss store hypotheses tried

Fig. 1. Number of three-trial blocks to solution for various shift groups in Experiment 1.

but rejected in short-term memory for a period of time until they are replaced by more recent information, it was expected that a shift to the Lag 1 H would be more difficult than one to the Lag 2 H, etc. The relative difficulty of a shift to an untried H should be a function of the time information is stored; on the basis of prior data (Erickson, 1968a) it was expected that this group would be roughly comparable to the Lag 3 group. These predictions were generally supported. With all Ss included, the Lag 1 and 2 groups were comparable, both having more difficulty with the shift problem than the Lag 3 and NH groups. The contrast (Lag 1 and Lag 2 vs Lag 3 and NH) approached significance ($F = 3.53$, $df = 1/81$; $p < .07$). With Ss who had three different hypotheses immediately prior to solving Problem 1, the linear trend in difficulty of the shift problem was significant ($F = 4.44$, $df = 1/43$; $p < .05$).

Examination of the data from the IDS and EDS groups showed that these two groups solved Problem 1 more quickly than Ss in the other groups (4.7. and 4.3 three-trial blocks to solution in Groups IDS and EDS, respectively, vs a mean of 10.3 blocks for the other groups). This is a result of the reasons Ss were assigned to these groups and is of little interest in and of itself. What is of interest is that the EDS group solved the shift problem faster than did the IDS group, 9.04 vs 14.33 blocks to solution ($t = 2.46$, $df = 46$, $p < .02$). This is the opposite of what is usually found with college-age Ss (see Wolff, 1967, for a review of this literature).

Table 1 shows summary statistics for some characteristics of hypothesis sampling for various groups of Ss. The probability of a win shift (an H change after a "correct" feedback) is high. The reasons for this are not entirely clear; substantially lower win-shift probabilities have been found using the same stimuli and essentially the same procedures (Erickson, 1968a). The major procedural difference between that study and the present is that three-trial blocks were used to infer Hs in the present study. Other studies, using longer blocks to infer Hs (Erickson, 1968a; Erickson & Block, 1969; Levine, 1966) have found that there are some response patterns not interpretable as Hs. The high win-shift probability found in the present study could be due to a combination of response patterns misinterpreted as Hs, accidental button-pressing errors, and "true" win shifts. It should be noted that IDS and EDS Ss, who solved Problem 1 very quickly, had a rather low win-shift probability, while Ss who solved Problem 1 but failed to solve their shift problem had a relatively high win-shift probability.

Table 1
Summary of Hypothesis Sampling Characteristics in Experiment 1

Statistic	Groups		
	Lag 1, 2 3. and NH	IDS and EDS	Ss Solving Only Problem 1
Problem 1			
Win-shift	.314	.144	.358
Lose-stay	.057	.079	.128
Local consistency	.815	.862	.664
IDS hypothesis	.112	.100	.114
Shift problem			
Win-shift	.374	.288	
Lose-stay	.094	.086	
Local consistency	.802	.766	
IDS hypothesis	.091	.086	

Note—Cell entries are probabilities.

For all groups, the lose-stay probability (failure to change the H following an "error" feedback) was low, on the order of that found in other experiments. Those Ss solving only Problem 1 had a somewhat higher lose-stay probability than the other Ss.

The probability of local consistency (choice of a new H that is consistent with information from the prior feedback trial) is consistent with values reported in other studies (Erickson & Block, 1969; Levine, 1969). Several recent theories have assumed local consistency to be one of the bases on which Ss choose Hs. If so, this probability should be close to 1.0; obviously, it is not. The Ss were able to make use of some information from the previous stimulus-feedback pairing but not all of it.

Local consistency was computed separately for new Hs after errors and after correct responses (lose shifts and win shifts) across both problems for all Ss. Local consistency on win-shift trials was .687, while on lose-shift trials it was .849. Win shifts may often be caused by S forgetting his current H when it is time for his next response. If this is true, a win-shift H might be chosen later than a lose-shift H (when the next stimulus is presented rather than during the intertrial interval) and Ss could have more trouble recalling the prior stimulus-feedback information for that reason.

Further analysis of the data showed that the probability of a new hypothesis being consistent with information from stimulus-feedback pairings farther back than the last one decreases to about .60 for outcome trials of Lag 2 and Lag 3, showing that Ss make relatively little use of more remote stimulus information.

The probability of an IDS hypothesis was quite low for all groups, substantiating the obtained difference between IDS and EDS shifts. When Ss received an "error" feedback, there was a strong tendency to test an H based on one of the other

dimensions rather than to test the reversal H on the same dimension.

EXPERIMENT 2

In Experiment 2, Ss received "error" feedback on each of the first three Hs tried, then were shifted to Lag 1, 2, or 3 H or to a NH for the remainder of the trials—until the problem was solved. Selection of Ss was stringent: Only those Ss who tried three different Hs during the first three "error" blocks were used; others were then given a series of "correct" feedbacks and allowed to think they had solved the problem; their data were not used.

A total of 88 male and female students enrolled in the introductory psychology course were run. Of these, 33 met the requirements of choosing three different Hs during the three "error" blocks and also solved the final problem. The apparatus and materials were the same as in Experiment 1.

The Ss were assigned randomly to four experimental groups, Lag 1, 2, or 3, or NH, as defined in Experiment 1. For example, the solution H for Ss assigned to the Lag 1 group was the H they tried during the third "error" block. After the final S selection, there were eight Ss in each group except Lag 2, which had nine.

The general procedure was the same as in Experiment 1. The Ss received thorough instructions and solved two practice problems; however, on the final problem, feedback was given every fourth trial instead of every third trial. Six sets of the 16 stimulus patterns were formed such that each set was divided into four four-trial blocks, each of which would permit the inference of an H. One of these sets was presented first to all Ss; the order of the other five sets was counterbalanced across Ss as much as possible. On the first three feedback trials, the feedback was always the opposite of S's response, indicating an "error." After these first three blocks, the feedback was consistent with the solution H.

Fig. 2. Number of four-trial blocks to criterion for various shift groups in Experiment 2.

Results

The first result to be noted is that it was rather difficult for Ss to meet the criteria for inclusion in the study. Of the 88 Ss run, only 33 had three different Hs during the three "error" feedback blocks and also solved their problem. Most Ss either repeated Hs or had response patterns uninterpretable as Hs during the three "error" blocks, and had to be rejected for purposes of this study. In addition, when the data from the 33 final Ss were examined, many non-H patterns were found during the solution phase of the experiment. During the precriterion period, after the "error" blocks and before the criterion run, 62 of the 234 response patterns were not interpretable as Hs, over 25%. This can be compared with 2%-5% non-H patterns reported by Erickson (1968a) and by Levine (1966). The three initial "error" blocks produced performance rather discrepant from that observed in other experiments. It should be noted that the initial three blocks contained sufficient information for a perfect information processor to eliminate all eight Hs logically. This presumably led to the numerous non-H blocks, even though Ss in these experiments are not perfect information processors (Levine, 1966; Erickson & Block, 1969).

Figure 2 shows the number of four-trial blocks to criterion for the various groups. The basic trends in the data were the same as those found in Experiment 1. Problem difficulty was quite linearly related to lag, Lag 1 being more difficult. The NH group approximated the Lag 3 group in ease of solution. These trends reached borderline

statistical significance. In an analysis using all four groups, the Lag 1 and 2 groups took more trials to solve the problem than did the Lag 3 and NH groups ($F = 3.12$, $df = 1/29$, $p < .10$). In the analysis using only the three lag groups, the linear trend approached significance ($F = 3.70$, $df = 1/22$, $p < .10$).

DISCUSSION

The basic results of Experiment 2 replicate those of Experiment 1, though they reached only borderline statistical significance. It seems that Ss remember Hs previously tried but rejected, but that these Hs remain in short-term memory only briefly. The approximately 2-H memory found in these studies is the same as that found previously (Erickson, 1968a; Erickson & Block, 1969), and seems to be a good estimate of H memory under the conditions of these experiments.

The data with respect to local consistency are rather interesting. Several recent models have assumed local consistency to be one of the bases on which Ss choose Hs to test; the data show rather clearly that the assumption is wrong in general.

Another aspect of the data that is extremely interesting is the fact that in Experiment 1 a reversal or IDS shift was found to be more difficult than a nonreversal or EDS shift for a group of college students. This result, the opposite of that usually found, has been replicated (Erickson, 1968b). With detailed instructions that stress the dimensions along which the stimuli can vary and point out the nature of the solution of the problem, Ss are likely to choose a new H which involves a different dimension from that just tested, rather than testing the reversal H along the same dimension. This is probably not true of Ss given the usual instructions in concept-shift experiments which tell S only that the stimuli fall in two classes and that his task is to learn how they are classified.

The model that seems most reasonable on the basis of this and other published evidence is a slight modification of those previously suggested. The S is assumed to solve simple concept-identification problems by selecting Hs to test. At the time a response is made, S is assumed to pay attention primarily to the dimension(s) involved in his current H, especially if he has made correct responses for several trials in a row. He may try to test more than one H at a time and pay attention to more than one dimension, but it is unlikely that he

attends actively to all dimensions of the stimuli.

On trials when he makes an error, he chooses a new H using whatever information is available to him at that time, including what he can remember about Hs previously tried and what he can remember about previous stimuli. To the extent that he cannot directly compare the prior stimulus and its feedback label, and has not attended to all of the dimensions of the stimulus when making his choice, his choice of a new H may not be perfectly consistent with the information from the prior trial.

Estimates of the amount of information available to S at the time of choice of a new H have been made by Erickson (1968a) and by Levine (1966, 1967, 1969). Current models for concept identification account fairly well for these estimates of the amount of information used at the time of H choice, but less well for exactly what mixture of prior Hs and prior stimulus classifications this information is.

REFERENCES

- BOWER, G., & TRABASSO, T. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford University Press, 1964. Pp. 32-94.
- ERICKSON, J. R. Hypothesis sampling in concept identification. *Journal of Experimental Psychology*, 1968a, 76, 12-18.
- ERICKSON, J. R. Problem shifts and hypothesis sampling in concept identification. Paper read at meetings of the Psychonomic Society, St. Louis, Missouri, 1968b.
- ERICKSON, J. R., & BLOCK, K. K. Concept learning with reinforcement of hypotheses and misinformative stimulus labeling. *Psychonomic Science*, 1969, 14, 183-184.
- 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.
- LEVINE, M. The size of the hypothesis set during discrimination learning. *Psychological Review*, 1967, 74, 428-430.
- LEVINE, M. Neo-noncontinuity theory. In G. Bower and J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 3. New York: Academic Press, 1969. Pp. 101-134.
- RESTLE, F. The selection of strategies in cue learning. *Psychological Review*, 1962, 69, 329-343.
- TRABASSO, T., & BOWER, G. H. *Attention in learning: Theory and research*. New York: Wiley, 1968.
- WOLFF, J. L. Concept-shift and discrimination-reversal learning in humans. *Psychological Bulletin*, 1967, 68, 369-408.