R E P O R T   R E S U M E S

ED 013 980   24   AA 000 248

RESPONSE-CONTINGENT POSTFEEDBACK INTERVALS IN CONCEPT IDENTIFICATION.

BY- BOURNE, LYLE E., JR.   AND OTHERS
WISCONSIN UNIV., MADISON

REPORT NUMBER BR-5-0216-TR-18   PUB DATE   FEB 67
CONTRACT OEC-5-10-154
EDRS PRICE   MF-$0.25   HC-$0.76   17P.

DESCRIPTORS- *LEARNING PROCESSES, *CONCEPT FORMATION, *FEEDBACK, *LEARNING THEORIES, *BEHAVIORAL SCIENCES, TIME FACTORS (LEARNING)

TWO EXPERIMENTS WERE CONDUCTED TO DETERMINE THE EFFECTS OF POSTFEEDBACK (POST-IF) INTERVALS OF DIFFERENT LENGTHS ON OVERALL PERFORMANCE IN CONCEPT IDENTIFICATION PROBLEMS. THE POST-IF INTERVAL REPRESENTS THE TIME BETWEEN PRESENTATION OF INFORMATIVE FEEDBACK ON ONE TRIAL AND THE ONSET OF THE NEXT STIMULUS PATTERN. THE EXPERIMENTS WERE DESIGNED TO TEST THE QUITE DIFFERENT IMPLICATIONS OF TWO THEORIES OF LEARNING--INCREMENTAL AND ALL-OR-NONE. FOUR CONDITIONS ON THE LENGTH OF THE POST-IF INTERVAL WERE USED--(1) SUB-OPTIMAL LENGTH ON ALL TRIALS, (2) SUB-OPTIMAL AFTER CORRECT RESPONSES, BUT NEAR OPTIMAL AFTER INCORRECT RESPONSES, (3) NEAR OPTIMAL AFTER CORRECT RESPONSES, BUT SUB-OPTIMAL AFTER INCORRECT RESPONSES, AND (4) NEAR OPTIMAL LENGTH ON ALL TRIALS. EXPERIMENT 1 CONSISTED OF 60 COLLEGE SUBJECTS WHO WERE SHOWN GEOMETRIC FIGURES THAT WERE VARIED IN SIX CHARACTERISTICS. EACH SUBJECT WAS TO CATEGORIZE THE FIGURES BY USING FOUR RESPONSE KEYS. ANALYSES OF TRIALS AND ERRORS SUPPORTED THE ASSUMPTION THAT LEARNING OCCURS ON ALL TRIALS AS SUGGESTED BY THE THEORIES BASED ON THE INCREMENTAL, ASSOCIATIVE PRINCIPLES. EXPERIMENT 2 USED 144 SUBJECTS WITH TWO RESPONSE KEYS, AND FOUND THAT THE RESULTS WERE SIMILAR BUT LESS CONSISTENT REGARDING EVERY-TRIAL LEARNING. IN BOTH EXPERIMENTS A LONGER POST-IF INTERVAL FACILITATED IDENTIFICATION OF CONCEPTS. THE AUTHOR CONCLUDED THAT THE RESULTS SUGGESTED THE EXISTENCE OF AT LEAST TWO TYPES OF BEHAVIORALLY DISTINCT SUBJECTS--(1) THOSE WHO ASSOCIATE STIMULUS PATTERNS OR ATTRIBUTES OF PATTERNS WITH CATEGORIES AND (2) THOSE WHO DELIBERATELY TEST HYPOTHESES. HE SUGGESTS THAT, BY PROPERLY ARRANGING A SERIES OF PRELIMINARY TASKS, SUBJECTS MIGHT BE PRETRAINED TO ASSOCIATE OR TO HYPOTHESIS-TEST. (AL)

RESPONSE-CONTINGENT POSTFEEDBACK INTERVALS IN CONCEPT IDENTIFICATION

RESEARCH AND DEVELOPMENT CENTER FOR LEARNING AND RE-EDUCATION

Technical Report No. 18

RESPONSE-CONTINGENT POSTFEEDBACK INTERVALS IN CONCEPT IDENTIFICATION

Lyle E. Bourne, Jr., David H. Dodd, Donald E. Guy

University of Colorado

and Donald R. Justesen

Veterans Administration Hospital, Kansas City

Research and Development Center
for Learning and Re-education
The University of Wisconsin
Madison, Wisconsin

February 1967

The writing of this paper was supported by a contract with the United States Office of Education, Department of Health, Education, and Welfare, under the provisions of the Cooperative Research Program.

Center No. C-03 / Contract OE 5-10-154

This research was undertaken in the Institute of Behavioral Science, University of Colorado, and the paper was written mainly during the summer, 1966, while the first author was a visiting scholar in the Research and Development Center.

The work was supported in part by the Grants MH 08315 and MH 11928-01 from the National Institute of Mental Health, United States Public Health Service, and by Grant GB 3404 from the National Science Foundation; by a National Science Foundation Cooperative Fellowship awarded to the second author; by a National Institute of Mental Health Research Fellowship MH 28,409-02 awarded to the third author; and by 8200 research funds provided by the United States Veterans Administration to the fourth author.

The authors are indebted to Nancy Wadsworth who collected most of the data.

PREFACE

One object of the R & D Center is to secure a more complete understanding of the processes involved in, and the conditions related to, the efficient acquisition, retention, and utilization of concepts. The present study clarifies two theoretical points of view, one of which holds that learning occurs on all trials in concept identification, the other that learning occurs only on the error trials. The study thus indirectly deals with the two highly significant theoretical positions in the psychology of learning: incremental learning or all-or-none. Evidence justifying an incremental point of view is presented in this paper; learning was inferred to have occurred on all trials rather than only on those on which errors were made.

The Center is fortunate indeed that Professor Bourne was affiliated with the Center as a visiting scholar during the summer of 1966. Although the data were gathered prior to the summer, the paper was written mainly during the summer.

Herbert J. Klausmeier
Co-Director for Research

CONTENTS

LIST OF TABLES

ABSTRACT

Slides containing geometric figures with six bi-valued dimensions were shown serially to 60 college Ss. Each S was to categorize the figures by pressing one of four response keys. A 4 x 3 design was employed with four conditions of response-dependent postfeedback interval, 1 or 15 seconds after a correct or an incorrect response, and three problems, each with two relevant dimensions. Analyses of trials and errors supported a theory which assumes that learning occurs on all trials, rather than one which specifies learning on error trials only. Analyses of response-by-feedback matrices showed that the relevant dimensions were not learned independently. A second experiment, using 144 Ss and two response keys, gave similar but less consistent results regarding every-trial learning. In both experiments a longer postfeedback interval facilitated identification of concepts.

I
INTRODUCTION

Increasing the length of the postfeedback (post-IF) interval, i.e., time between presentation of informative feedback on one trial and the onset of the next stimulus pattern, facilitates performance in concept identification tasks. It is known, moreover, that the optimal interval increases directly with task complexity (Bourne & Bunderson, 1963; Bourne, Guy, Dodd & Justesen, 1965). These data imply the occurrence of significant problem-solving activities during the interval. Theories of concept learning based either on incremental, associative principles (e.g., Bourne & Restle, 1959) or on the assumption of all-or-none learning (e.g., Bower & Trabasso, 1963) give plausible and consistent descriptions of the data. The facilitative effects could be ascribed to the greater opportunity provided by a longer interval for associating relevant cues with responses (and for adapting irrelevant cues) or for more adequate sampling, in some sense, from the pool of possible solutions (Restle &Emmerich, 1966).

While both theoretical positions easily accommodate the available data, the interpretations of the two models are different formally and embody contrasting implications subject to further empirical test. Learning, within the hypothesis-selection framework, is a response-dependent process. Resampling from the hypothesis pool, and therefore attainment of problem solution, occurs (by assumption) only on error trials. The effectiveness of variation in the post-IF interval must be limited, then, to trials on which S makes an incorrect category response. On the other hand, cue-conditioning theory implies that learning occurs on all trials. Therefore, the influence of interval length is indifferent to the correctness of S's response.

The present experiments were designed as a simple test of these implications. Four conditions on the length of the post-IF interval were arranged: The interval was fixed at (I) sub-optimal length on all trials; (II) sub-optimal after correct responses, but near optimal after incorrect responses; (III) near optimal after correct responses, but sub-optimal after incorrect responses; and (IV) near optimal length on all trials. Any model which embodies the notion that learning occurs only on error trials, and (therefore) that longer post-IF periods are beneficial only on error trials, implies the following performance ordering among conditions from worst to best: I = III > II = IV. The ordering to be expected if learning occurs on all trials is: I > II = III > IV. The purpose of these experiments is to determine which, if either, of these orders more accurately reflects overall performance in concept identification problems.

II
EXPERIMENT I

METHOD

Task and Apparatus

Each S was given a detailed preliminary description of the stimulus material, the type of responses required, the meaning of informative feedback signals and the general nature of the solution to be attained. Several stimulus patterns (geometrical designs) were shown to S on a translucent viewing screen set in an opaque partition. The patterns varied in six binary dimensions—size (large-small), color (red-green), vertical position (top-bottom), horizontal position (left-right), form (square-triangle), and number (one-two figures)—each of which was described and illustrated for S.

The Ss were told that (a) the task was to learn how to sort the patterns into four categories, (b) the solution was based on two relevant or critical stimulus dimensions, and (c) the four conjunctive combinations of the two values on these two dimensions defined the categories. Several sample solutions, using randomly selected pairs of dimensions, were described. Following these instructions, Ss solved (with hints and guidance provided by E, if necessary) a pretraining problem, with horizontal and vertical position as the relevant dimensions. Subjects were given an opportunity to ask any remaining questions after which the experimental problem was begun. Subjects were required to achieve a criterion of 16 correct responses in a row.

Stimulus patterns were prepared as 2 x 2 inch slides. They were presented serially, projected to the rear of the translucent viewing screen. The S responded to each pattern, after a self-paced inspection interval, by pressing one of four buttons mounted in a control panel below the screen. Immediately after each response, the pattern was removed—the screen becoming blank—and a feedback signal lamp was turned on for 1 second over the correct response button. The time between the termination of the feedback signal and the appearance of a new stimulus pattern, i.e., the post-IF interval, was 1 second for all Ss during the pretraining problem and either 1 (sub-optimal) or 15 (near optimal) seconds (Bourne et al., 1965) in the experimental problem depending on the condition to which S was assigned. The presentation of stimulus patterns and feedback and the timing and recording functions were controlled electronically with an apparatus described elsewhere (Bourne & Haygood, 1959).

Subjects and Design

The Ss were 60 undergraduate college students who volunteered for participation. They were assigned randomly but in equal number to experimental treatments prescribed by a 4 x 3 factorial design. The post-IF conditions were the four possible combinations of 1 and 15 second intervals after correct and incorrect responses. The second variable, problems, was determined by the particular pair of stimulus dimensions chosen as relevant to solution: (A) size and color, (B) size and form, (C) color and form. All problems contained two relevant and four irrelevant dimensions.

RESULTS

Main Analyses

Mean numbers of trials to last error were 91.6, 52.8, 48.5, and 43.8 for post-IF conditions I (1-1), II (1-15), III (15-1), and IV (15-15), where the two numerals in parentheses indicate interval length after correct and incorrect responses, respectively. Corresponding mean errors to solutions were 58.2, 31.0, 29.5, and 25.4. Analyses of variance show the overall difference among post-IF conditions to be reliable for both trials, $F(3,48) = 3.15$, and errors, $F(3,48) = 4.16$,

$p < .05$. The difference between groups with 1 (Conditions I and II) and 15 (Conditions III and IV) second intervals after correct responses was significant, $F(1,48) = 4.64$ and 5.86, $p < .05$, for trials and errors, respectively; however, the difference between 1 (Conditions I and III) and 15 (Conditions II and IV) second intervals after incorrect responses was not, $F(1,48) = 2.73$ and 3.80, $p > .05$. Problems and the interaction of problems and post-IF conditions were insignificant sources of variance.

A test for ordered hypotheses (Page, 1963) was used to evaluate theoretically-derived expectations. This analysis provides significant support for an ordering implied by theories which assume learning on all trials, I > II = III > IV, $L = 572.5$, $p < .001$, but not for the analogous ordering based on the notion that learning occurs only on error-trials, I = III > II = IV, $L = 189$, $p > .10$.

Supplementary Analyses

It is impossible to determine the particular trials, if any, on which S selects or identifies the first of the two relevant dimensions; thus, the usual techniques for determining stationarity in presolution response probabilities, i.e., backward, stationarity and Vincentized learning curves and their associated statistics, are of doubtful applicability to four-category problems. It might be noted, however, that incorrect response probabilities computed for all Ss on the last, the last 5, and the last 10 trials before solution (excluding the final error trial) were .31, .39, and .44, respectively. Binomial tests indicate that these values are all significantly less than .5 ($z = 2.97$, 3.87, and 2.42, respectively, $p < .05$).

Each trial for any S can be represented as an entry in a response-by-feedback (confusion) matrix. Table 1 shows the Vincentized fre-

Table 1

Response by Feedback (Confusion) Matrices for Each of Four Performance Quartiles, Experiment I

		S's Response 1	2	3	4	Σ
Quartile 1	Feedback 1	72	67	38	48	225
	2	56	68	30	45	199
	3	55	29	79	41	204
	4	43	69	50	64	226
	Σ	226	233	197	198	854 = ΣΣ
Quartile 2	Feedback 1	83	58	47	31	219
	2	52	63	32	66	213
	3	45	34	90	44	213
	4	30	48	55	74	207
	Σ	210	203	224	215	852 = ΣΣ
Quartile 3	Feedback 1	76	67	41	25	209
	2	41	93	33	45	212
	3	44	31	87	54	216
	4	20	49	58	84	211
	Σ	181	240	219	208	848 = ΣΣ
Quartile 4	Feedback 1	103	42	38	32	215
	2	55	99	16	42	212
	3	24	28	108	43	203
	4	24	51	51	95	221
	Σ	206	220	213	212	851 = ΣΣ

quencies with which each of the four responses and each of the four feedback signals occurred together in each of the four quartiles. Several results are notable. Clearly, no matrix is homogeneous. The proportion of correct responses is .33, .36, .40, and .48 in quartiles 1-4, respectively. Even in the first quartile, this is a significant departure from chance (.25) responding ($z = 5.43$, $p < .01$). Further, the probabilities of making a correct response to either relevant dimension (on either subproblem [Trabasso & Bower, 1964]) significantly exceed .5 in all quartiles ($z \geq 4.28$, throughout, $p < .01$).

A test was made of an assumption common to cue-conditioning (Bourne & Restle, 1959) and certain hypothesis-testing models (Trabasso & Bower, 1964) that Ss learn the two subproblems independently. If this independence assumption is tenable, the probability of correct responding on one subproblem (A) should be independent of the probability of being correct on the other (B). Thus the conditional probability of being correct on A given wrong on B—$P(R_A/W_B)$—should estimate the probability of being correct on A—$P(R_A)$. And the product $P(R_A/W_B) \times P(R_B/W_A)$ should estimate the probability of being correct on both dimensions—$P(R_{AB})$—which is the usual definition of a "correct" response in the four-category problem. Table 2 presents the comparison of the predicted and obtained proportion of correct responses. These values are similar except in the fourth quartile where the obtained value is significantly greater than the predicted ($z = 4.55$, $p < .01$).

Table 2

Obtained vs. Predicted Proportions of Correct Responses by Quartiles, Experiment I. Based on the Assumption of Independence of Subproblems

	Quartile			
	1	2	3	4
Predicted	.341	.385	.416	.399
Obtained	.331	.364	.401	.476

The data of the fourth quartile were further partitioned into the four post-IF conditions of the experiment and the independence test repeated on each of these. The results, shown in Table 3, indicate that both Conditions II and IV (15 second delay after an error) show significant departure of obtained values from predicted ($p < .01$), while Conditions I and III show small departures, though in the same direction.

Table 3

Obtained vs. Predicted Proportions of Correct Responses by Conditions in Fourth Quartile, Experiment I. Based on the Assumption of Independence of Subproblems

	Condition			
	I	II	III	IV
Predicted	.424	.322	.389	.300
Obtained	.451	.513	.458	.503
Z, differences	.95	5.48	1.76	5.51

Tests of goodness of fit to a binomial distribution were performed on the 5 trials preceding the last error. The obtained and theoretical frequencies of correct responses on these trials are presented in Table 4. Expectations were computed on two bases: (a) by assuming that all Ss have acquired one subproblem but not the other and are thus responding correctly 50% of the time (the maximal percentage consistent with hypothesis-testing notions in this type of problem), and (b) using the obtained mean proportion of correct responses (.613) over these trials. The chi-squares for goodness of fit are both significant ($p < .01$), being 8.02 and 13.25 for .5 and .613 distributions, respectively.

SUPPLEMENTARY DATA

For Conditions II and III of the preceding experiment, the post-IF interval took values of 1 or 15 seconds, depending on the correctness of S's response. The actual mean interval for these Ss was 7.8 seconds. As a means of further evaluating the equivalence of intervals after errors and correct responses, an independent group of 18 Ss was given a four-category task with 7.8 seconds post-

IF duration on all trials. In all respects, task conditions were identical with those of Experiment I, including the random assignment of Ss to the three different problems. Mean number of trials and errors to criterion were 52.3 and 33.7, respectively. The mean probabilities of error over the last, the last 5, and the last 10 presolution trials were .29, .41, and .45, respectively. As such, the data are consistent with Experiment I and tend to support the notion that post-IF interval length is indifferent to the correctness of S's response.

Table 4

Distribution of Observed vs. Binomial Frequencies of Correct Responses on Five Trials Preceding Last Error

	Number of Correct Responses Prior to last error					
	0	1	2	3	4	5
Observed Number of Ss	1	9	11	16	13	10
Binomial (.5,5)	1.88	9.38	18.75	18.75	9.38	1.88
Binomial (.613,5)	.54	4.02	13.08	20.70	16.38	5.16

III
EXPERIMENT II

Most of the results which are compatible with the all-or-none, error-trial learning principle have been obtained in simple, two-response concept identification problems. In view of Experiment I, it might be argued that task complexity is an important factor controlling the nature of changes in response probabilities and the opportunity for learning to occur. Experiment II, having essentially the same design as the first study, was conducted to determine whether performance in four conditions of post-IF interval would have a different alignment in a simple two-response task.

METHOD

In all respects, except those mentioned below, the procedure was the same as that used in Experiment I.

The study was planned as a 4 (post-IF conditions) by 3 (problems) factorial, with 4 Ss to be randomly assigned to each of the 12 cells. In fact, however, this plan was replicated three times with minor procedural modifications so that, overall, the study can be described as a 4 x 3 x 3 (replications) design. In total 144 Ss served in the experiment, 4 in each of the 36 cells.

Replication A: After brief instructions about the task, each S solved a problem with one relevant dimension (either size, color, or form) and five irrelevant dimensions. Post-IF intervals were 1 or 15 seconds, depending on the condition to which S was assigned. Performance differences among the post-IF conditions were small and unreliable, a result which might reflect the fact that 15 second intervals are supra-optimal in simple (two-response) problems (Bourne et al., 1965). Replication B: In the second replication the longer interval was reduced to 9 seconds. The trend across conditions was essentially the same as those in Experiment I and Replication A, but statistically unreliable. Replication C: The study was repeated with the addition of a preliminary practice problem. The attempt was to reduce inter-S variability by providing a clearer understanding of the form of solution to be attained. The practice problem was the same for all Ss, one relevant and two irrelevant dimensions and a constant (1 second) post-IF interval.

COMBINED RESULTS

Main Analyses

While the trends across conditions were roughly the same, the differences among them within each replication were unreliable. Strictly speaking, the results are inconclusive and might be taken to imply only that the length of post-IF interval and/or the contingency between the interval and S's response are insignificant conditions in two-response problems. An analysis of data combined over the three replications, however, helps to clarify the outcome.

Mean trials to last error for Conditions I, II, III, and IV were 19.3, 12.1, 11.8, and 11.4, respectively. Corresponding mean errors were 8.7, 6.1, 6.4, and 6.2. Statistically, the overall difference in trials was insignificant, $F(3, 108) = 1.89$, as were the differences between short (Conditions I and II) and long (Conditions III and IV) intervals after correct responses and between short (Conditions I and III) and long (Conditions II and IV) intervals after incorrect responses. The contrast reflecting some (Conditions II, III, and IV) versus no (Condition I) long intervals was, however, large and reliable, $F(1, 108) = 5.65$, $p < .01$. The alignment of conditions is notably like that observed in Experiment I, which adds some confidence in the importance of this difference. Incorrect response probabilities, computed for all Ss were .37, .44, and .44 on the trial, the 5 trials, and the 10 trials before the last error, all of which are reliably less than .5, z = 2.62, 2.57, and 3.30, respectively, $p \leq .01$.

Differences among the three problems were significant, F (2,108) = 5.65, $p < .01$; however, the source of variance identified with replications and all the interaction terms were negligible.

Supplementary Analyses

As in Experiment I, goodness of fit tests to a binomial distribution were performed using the 5 trials preceding the last error. Chi-squares using the theoretical (.5) and obtained (.44) probability of incorrect response were nonsignificant.

The failure to show significant differences among post-IF conditions could reflect an artificial ceiling effect produced by the simplicity of the solution required. Analysis of the stimulus sequences revealed that sufficient information for solution was available at the end of the fifth trial for each of the three problems used. Of 144 Ss, 73 (51%) solved in five or fewer trials; thus, the median number of trials is also 5. As shown in Table 5, Ss solving at or below the median (five or fewer trials) are evenly distributed across the post-IF interval conditions, $\chi^2(3) = .77$, p .50.

Considering only the "poorer" Ss, i.e., those who required six or more trials to criterion, the means were 37.65, 20.10, 21.78, and 22.19 trials for Conditions I, II, III, and IV, respectively. The overall difference among conditions was significant, F (3,67) = 3.49, p .05. Differences between short (Conditions I and II) and long (Conditions III and IV) intervals after correct responses were insignificant, F (1,67) = 2.03, as were differences between short (Conditions I and III) and long (Conditions II and IV) intervals after incorrect responses, F (1,67) = 3.79. The interaction of these variables, however, was reliable, F (1,67) = 4.66, $p < .05$, indicating that optimal intervals occurring on either trial-type are sufficient to produce maximal performance.

Table 5

Number of Ss Taking More or Fewer Than Five Trials to Solve, Experiment II.

	Condition			
	I	II	III	IV
Below median (0-5 trials)	19	16	18	20
Above median (6 or more trials)	17	20	18	16

IV
DISCUSSION

OVERALL TRENDS

The outcome of these experiments is clearly consistent with the notion that there are significant behavioral changes (learning) on all trials rather than on error trials only. Performance is better when there are some as contrasted with no optimal post-IF intervals, and better still when the interval on all trials is long (at least in four-response problems). Moreover, it does not appear to make any reliable difference whether the longer interval is limited to correct or incorrect response trials. While these findings are not inconsistent with the notion that learning is a response-dependent process (i.e., what S learns or how behavior is changed might differ between correct and incorrect response trials), they clearly conflict with an error-trial-only learning principle. Similar conclusions have been drawn by Suppes & Schlag-Rey (1965) and by Levine (1966).

The magnitude of differences in trials and errors to solution among Conditions II-IV (for which some or all trials are accompanied by long post-IF intervals) is small in both experiments. Having some (approximately 50%, Conditions II and III) is nearly as facilitative as having all (Condition IV) long intervals under the performance conditions used here. There is then the possibility of a ceiling effect, which might be explored by varying the percentage of long intervals (randomly assigned to trials, irrespective of S's response) across conditions in these problems. On the basis of current data, performance would be expected to improve directly and monotonically with this percentage up to some optimum.

INDEPENDENCE OF SUBPROBLEMS

Two particular theoretical developments, one adopting the every-trial learning principle (Bourne & Restle, 1959) and the other based on the error-trial-only principle (Bower & Trabasso, 1963) embody the assumption that the two relevant dimensions of a four-category task are learned independently, as subproblems. Analyses on the data of Experiment I, reported in Tables 2 and 3, are clearly inconsistent with this assumption. Responses with respect to the two dimensions are dependent, though the behavioral character of this dependence is uncertain. Trabasso & Bower (1964) entertain the possibility of a paired associate stage, during which S associates each response category with a conjunction of attributes, subsequent to the dimension selection stage in four-category learning. The paired associate stage could imply "subproblem" dependence during the later trials of learning. However, the partitioning of the fourth quartile confusion matrix into the four post-IF conditions raises additional questions. This analysis indicates that there is greater subproblem dependence in Conditions II and IV, where the longer post-IF intervals follow errors, than in Conditions I and III. If dependence reflects paired associate learning then a different result would be expected, for distribution of practice accelerates the process, if anything (Underwood, 1961). Actually, there is no obvious interpretation for the pattern revealed in the fourth quartile confusion matrices. Further work is needed to determine its replicability. If replicable, however, the results demonstrate that at least under some conditions the principle of independent subproblems is an inappropriate description of S's behavior.

LEARNING DURING THE TERMINAL TRIALS

Results of the analysis of response probabilities during the last five trials of Experiment I indicate a significant departure from the binomial distribution which would be expected on the independent events assumption of some models (e.g., Bower & Trabasso, 1963). The result might be attributed to incrementally changing response probabilities, but this would imply a distribution more peaked (leptokurtic) than the binomial. Instead, the obtained distribution is flatter—an outcome which probably can be traced to the considerable variability in response probabilities among Ss (rather than across trials). Individual differences remain a serious challenge to current mathematical descriptions of learning. The present result might be taken speculatively to suggest the existence of at least two types of behaviorally distinct Ss, those who associate stimulus patterns (or attributes thereof) with categories and those who deliberately test hypotheses (see also Osler & Fivel, 1961). During the last five trials, hypothesis-testers presumably would have correct response probabilities of .5 (or .25) while the corresponding probability for a gradual learner would be approaching 1.0. Summing the scores of these two types of Ss would lead to the kind of distribution obtained. Speculative though such a description is, it seems possible to pretrain Ss to associate or to hypothesis-test by properly arranging a series of preliminary tasks. Analysis of the performance of these S-types, singly and combined, might clarify some of the features of the data reported here.

REFERENCES

Bourne, L. E., Jr., & Bunderson, C. F. Effects of delay of information feedback and length of postfeedback interval on concept identification. Journal of Experimental Psychology, 1963, 65, 1-5.

Bourne, L. E., Jr., Guy, D. E., Dodd, D. H., & Justesen, D. R. Concept identification: The effects of varying length and informational components of the intertrial interval. Journal of Experimental Psychology, 1965, 69, 624-629.

Bourne, L. E., Jr., & Haygood, R. C. The role of stimulus redundancy in concept identification. Journal of Experimental Psychology, 1959, 58, 232-238.

Bourne, L. E., Jr., & Restle, F. Mathematical theory of concept identification. Psychological Review, 1959, 66, 278-296.

Bower, G., & Trabasso, T. Reversals prior to solution in concept identification. Journal of Experimental Psychology, 1963, 66, 409-418.

Levine, M. Hypothesis behavior by humans during discrimination learning. Journal of Experimental Psychology, 1966, 71, 331-338.

Osler, S. F., & Fivel, M. W. Concept attainment: I. The role of age and intelligence in concept attainment by induction. Journal of Experimental Psychology, 1961, 62, 1-8.

Page, E. B. Ordered hypotheses for multiple treatments: A significant test for linear ranks. Journal of the American Statistical Association, 1963, 58, 216-230.

Restle, F., & Emmerich, D. Memory in concept attainment: Effects of giving several problems concurrently. Journal of Experimental Psychology, 1966, 71, 794-799.

Suppes, P., & Schlag-Rey, M. Observable changes of hypothesis under positive reinforcement. Science, 1965, 148, 661-662.

Trabasso, T., & Bower, G. Component learning in the four-category concept problem. Journal of Mathematical Psychology, 1964, 1, 143-169.

Underwood, B. J. Ten years of massed practice on distributed practice. Psychological Review, 1961, 68, 229-247.