A COMPARISON OF THREE MODELS OF LEARNING WITH PROBABILISTIC CUES

by

STEVEN PATRIC ROGERS

A.B., Stanford University, 1964

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Psychology

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1966

Approved by:

Robert C. Haygood
Major Professor

ACKNOWLEDGEMENTS

To. . .

My parents, who taught me the difference between school and education

Alex Bavelas, who made psychology exciting

Robert C. Haygood, without whose advice, encouragement, and assistance this work would have been impossible

The National Aeronautics and Space Administration, who provided my financial support over the past two years

Omar the tent-maker, whose guidance shall not be forgotten

. . . this thesis is very gratefully dedicated.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

LIST OF TABLES iv

LIST OF FIGURES vi

CHAPTER

1 INTRODUCTION 1

2 EXPERIMENT 1; MISINFORMATIVE FEEDBACK 9

3 EXPERIMENT 2; NONCONTINGENT FEEDBACK 20

4 EXPERIMENT 3; OCCURRENCE PREDICTION 27

5 EXPERIMENTS 1, 2, AND 3; COMBINED ANALYSES 33

6 DISCUSSION 44

7 SUMMARY 52

REFERENCES 55

APPENDIX 58

LIST OF TABLES

Table Page

1. Analysis of Variance Summary Table for Absolute Errors in Experiment 1 (cue-contingent situation) 13

2. Analysis of Variance Summary Table for Absolute Errors in the Last Block of Ten Trials of Experiment 1 13

3. Analysis of Variance Summary Table for Response Latencies in Experiment 1 (cue-contingent situation) 16

4. Analysis of Variance Summary Table for Response Latencies in the First Block of Forty Trials in Experiment 1 . . . 16

5. Analysis of Variance Summary Table for Major Hypothesis Changes by Subjects in Experiment 1 17

6. Analysis of Variance Summary Table for Minor Hypothesis Amendments by Subjects in Experiment 1 17

7. Analysis of Variance Summary Table for Total Alterations of Hypotheses (changes and amendments combined) in Experiment 1 18

8. Analysis of Variance Summary Table for Response Latencies in Experiment 2 (cue-noncontingent situation) 22

9. Analysis of Variance Summary Table for Response Latencies in the First Block of Forty Trials in Experiment 2 . . . 22

10. Analysis of Variance Summary Table for Major Hypothesis Changes by Subjects in Experiment 2 23

11. Analysis of Variance Summary Table for Minor Hypothesis Amendments by Subjects in Experiment 2 23

12. Analysis of Variance Summary Table for Total Alterations of Hypotheses (changes and amendments combined) in Experiment 2 24

13. Analysis of Variance Summary Table for Absolute Errors in Experiment 3 (cueless situation) 29

14. Analysis of Variance Summary Table for Response Latencies in Experiment 3 (cueless situation) 29

LIST OF TABLES (continued)

Table Page

15. Analysis of Variance Summary Table for Response Latencies in the First Block of Forty Trials in Experiment 3 . . . 32

16. Analysis of Variance Summary Table for Absolute Errors Based on Experiments 1 and 3 Combined (cue and cueless) . 35

17. Analysis of Variance Summary Table for Response Latencies in Experiments 1, 2, 3 Combined 37

18. Analysis of Variance Summary Table for Major Hypothesis Changes by Subjects in Experiments 1 and 2 Combined . . . 40

19. Analysis of Variance Summary Table for Minor Hypothesis Amendments by Subjects in Experiments 1 and 2 Combined . 41

20. Analysis of Variance Summary Table for Total Alterations of Hypotheses (changes and amendments combined) in Experiments 1 and 2 42

LIST OF FIGURES

Figure Page

1. Absolute error over trials by per cent MF, Exp. 1 14

2. Hypothesis modifications over trials, Exp. 1. 19

3. Hypothesis modifications over trials, Exp. 2. 25

4. Absolute error over trials by per cent Not-X events, Exp. 3 30

5. The interaction of per cent Not-X events and the after-error, after-correct latency effect, Exp. 3 31

6. The interaction of 40-trial blocks and the after-error, after-correct latency effect, Exp.3 33

7. The effects of groups, trials, and their interaction on absolute error, Exps. 1 and 3 combined 36

8. The interaction effect of 40-trial blocks and per cent MF, MC errors, and Not-X events upon latency 39

9. The effects of groups, trials, and their interaction upon hypothesis modifications, Exps. 1 and 2 combined 43

CHAPTER 1

INTRODUCTION

In traditional learning and problem-solving studies, there is typically an element of homogeneity in the experimental work; S's task is soluble, and his errors are corrected. Clearly this is not the case in many "real world" learning situations. Man is not always presented with immediate, complete, and accurate information concerning the appropriateness of his actions. Instead, he is perenially "forecasting" the probable results of his behavior, or that of others, and many of his responses must be learned with inconsistent consequences, and intermittent rewards. Much of his feedback is defective in one form or another. Three primary experimental paradigms have been used to study this kind of learning, all falling under the general classification of probability learning.

Three Models of Probability Learning

Occurrence prediction. Many experimenters have asked S to predict the occurrence of an event, for example, the illumination of one of two small lights, with absolutely no physical attributes or other cues available to use as aids in prediction. The independent variable in this model is the proportion of the two events in an otherwise random sequence of events.

Misinformative feedback. A frequently used paradigm is one in which the optimal response is made by S, but is followed by consequences indicating an error, or a nonoptimal response is followed by information signaling correct action. The S is thus misinformed about the correctness of his response, or, in other words, receives misinformative feedback (MF). In this case, of course, there are stimulus cues with which to make judgements, and often more independent variables than the proportion of MF.

Noncontingent feedback. This form of experiment, in its complete sense, has received little attention. It requires that the reinforcement be totally unrelated to the actions or predictive constructs of S. Reward simply occurs in a random, or perhaps fixed ratio manner. Spurious cues are presented to S, and proportion of noncontingent reinforcement is the major independent variable.

Research With Occurrence Prediction

The earliest of the occurrence-prediction studies was that of Humphreys (1939), who used a technique of verbal expectations regarding the occurrences of the second of two successive stimuli, i.e., lights on a panel. His first group was given 25 trials with the left light always following the right, and rose rapidly toward 100 per cent positive expectation. The second group, given 50 per cent (random) occurrences of the left following the right light, continued to vary around 50 per cent affirmative predictions.

Grant, Hake, and Hornseth (1951) improved upon this experiment by including reinforcement percentages of 25 per cent, and 75 per cent. At the end of the training series, each of the four groups was found to be responding in a proportion roughly equivalent to the percentage of reinforced trials. This behavior has subsequently been called probability matching (e.g., Estes, 1964). The Ss in the 25 and 75 per cent groups could have, but did not maximize their probabilities of correct prediction by responding in accord with the more likely alternative 100 per cent of the time. Other expectancy studies, using various proportions of reinforcement, but with similar results, have been performed by Estes and Straughan (1954); Friedman, Burke, Cole, Keller, Millward and Estes (1964); Graeno (1962); and Messick and Solley (1957).

Research With Misinformative Feedback

It is noteworthy that both Humphreys and Grant used cueless situations. If instead, S is flooded with cues, does probability matching still occur? To investigate this question, it is necessary to turn to MF. By replacing the signal lights of the cueless situation with more complex stimuli in which relevant attributes are related to reinforcement on some percentage basis, an MF situation is created.

Goodnow and Postman (1955), using geometrical designs in a two-choice problem-solving context, included MF percentages of 0, 10, 20, 30, 40, and 50. The Ss responded with probability matching behavior even though they did not realize that the task was a probability situation, and were trying to find lawful solutions to the problem. Morin (1955), in a rather similar experimental situation, found maximizing behavior in groups prewarned of MF, but a lack of probability matching in the unwarned groups. Morin, however, did not elaborate on this striking finding. Pishkin (1960), used an MF design with 0, 10, 20, 30, and 40 per cent MF, with problem complexities of 1, 3, and 5 irrelevant dimensions. He found that asymptotic mean error scores did not differ significantly from mean values predicted by a probability-matching hypothesis. This was true over all complexity groups, and in the 10, 20, and 30 per cent MF groups. Pishkin's study has been replicated and extended by Johannsen (1962), who used 0, 12.5, 25, and 37.5 per cent MF with 1, 3, and 6 irrelevant dimensions.

Research With Noncontingent Feedback

A noncontingent (NC) feedback study attempts to determine what amount of the process of developing and testing hypotheses is purely a function to the occurrence of reward, without confounding this with the specific stimulus or response aspects of the situation preceding reinforcement. This may be accomplished by use of a predetermined reinforcement schedule which is unrelated to any stimulus attribute, or any specific verbal or motor response of S.

Jenkins (1959) made use of a mathematical concept formation task with NC positive reinforcement in percentages ranging from 33 1/3 to 100. His Ss offered extremely complex hypotheses even though they were told that it might be impossible to solve the problems. There was a tendency toward less complexity with high NC positive feedback, although this did not reach significance. New hypotheses were introduced at the greatest rate early in learning. It was also noticed that a variety of hypotheses were introduced by S, and held simultaneously, though he might have received no errors.

Wright (1962) extended this research by using 16 possible responses, 325 trials, and 0, 20, 50, 80, and 100 per cent NC positive feedback.

His results indicated that hypothesis complexity is greatest at 50 per cent, and that probability of response repetition was a linear function of proportion of NC positive feedback. Although the response levels were below what would be expected in probability matching behavior, this may have resulted from the 16, as opposed to the usual 2 possible responses. Generally Ss do not detect that feedback is noncontingent. There is recent evidence (Jenkins and Ward, 1965), that it is very difficult, if not impossible, for them to do so.

Thus the general finding is that under conditions of random NC reinforcement, Ss respond in an orderly, systematic fashion, although their hypotheses vary greatly both within and between Ss. One is reminded of Skinner's (1948) development of "superstition" in pigeons by using fixed interval reinforcement schedules, and of similar results found by Guthrie (1946), in examining the behavior of cats in puzzle-boxes. It is probable that the reinforcement of unrequired responses, on nonrequisite particulars of a response occurs frequently in learning experiments, and certainly even more often in real-life situations.

Contentions Regarding Probability Matching

Exactly why probability matching occurs remains unclear. Bruner, Goodnow, and Austin (1956) submit that it is probably the result of a combination of three factors: (a) the hope of a unique solution, (b) interest in the less probable alternative, and (c) the need for a direct test of a hypothesis. There is some experimental support for this position (e.g., Goodnow, 1955a, 1955b; Morin, 1955; Hyman and Jenkin, 1956), but none of it explains why the response levels should arrive at probability matching, rather than some other level, excluding 100 per cent.

It is Bourne's contention (1963) that probability matching represents only a transitory level of performance. Thus his Ss were run for a total of 1200 trials, by the end of which nearly all Ss adopted the maximizing, 100 per cent response behavior. It could, however, be argued that after such a protracted period of failure, Ss simply "gave up" and no longer attempted to solve the problem.

Goodnow and Pettigrew (1955) suggest that a "win-stay, lose-shift" strategy must inevitably result in probability matching. Hake and Hyman (1953) point out that probability matching may result from following a sequential rule based upon the occurrences of prior events. Estes (1964) warns that apparent probability matching in some cases may be the result of a methodological artifact in grouped data, with some Ss rising rapidly toward 100 per cent, and others varying around 50 per cent response levels. Along with specific strategies, "subjects are using the overall relative frequency of the two events as a general guide" according to Bruner, et al. (1956, p. 188). To what extent Ss are aware of the actual percentages, however, is not clear from the literature.

Behavioral Characteristics of Problem Solving; Mathematical Models

Recently there have been several attempts to construct mathematical models to describe error and trials data in concept-formation situations. These models have enjoyed fair success in describing behavior in such instances. Both "stimulus sampling theory," as outlined by Atkinson, Bower, and Crothers (1965), and Restle's "hypothesis sampling theory" (1961), make two basic assumptions regarding Ss' behavior. First, if the response to a hypothesis sampled on trial n was correct, the same hypothesis will be maintained on trial n+1. Second, if the response is erroneous, S randomly chooses from the pool of possible hypotheses (resampling with replacement), and obtains a new hypothesis for trial n+1. Some experimental support for these assumptions is provided by Bourne (1965), and Levine (1966). These mathematical models are designed primarily for problems which have some strategies yielding 100 per cent correct feedback. If, however, S believes that he may completely solve the problem by some means, the behavioral assumptions should not differ in the probability-learning task.

Thus one is led to two clear predictions. First, after a correct response S will maintain the hypothesis which provided him with this success, but after an error, S will either change his hypothesis completely or modify it to some extent. Second, if S chooses a new hypothesis, it must be expected that some measurable amount of time is required for him to do so, and that this time will be added to the temporal delay typical in choice behavior. That is, response latency after an error should exceed that after a correct response. Such an effect has been found by Erickson, Zajkowski, and Ehmann (1965) in a standard concept formation task, but has not been examined in a probability-learning situation.

Plan of the Present Study

From the above review of the history, theoretical stances, and general commentary upon probability matching and behavioral characteristics of problem-solving, numerous questions arise. The goal of the present study is to critically examine but a few of them, by the use of three related experiments, and comparisons among them when applicable. Collection of error data, supplemented by recordings of Ss' verbalized hypotheses in these three experimental paradigms are used to explore the origin and generality of probability matching behavior. The hypothesis data are coupled with measurements of response latencies in order to investigate conceptual-behavioral characteristics of problem-solving, particularly the assumptions of the mathematical models. Observations are made pertaining to the course of probability learning in the three models in order to determine how it may differ as a result of the particular qualities of each situation.

Experiment 1 is a standard MF situation except that the number of irrelevant dimensions is limited only by S's imagination. Experiment 2 is a truly noncontingent paradigm; no matter what hypothesis is used by S, or what response he gives, his feedback of "right" or "wrong" is predetermined. Experiment 3 is a standard occurrence-prediction study, not unlike that of Humphreys. The Ss in all three experiments are told that the problem is soluble. It does not appear that these three important models for the study of probability learning have ever before been compared in the same study.

CHAPTER 2

EXPERIMENT 1

MISINFORMATIVE FEEDBACK

Method

Subjects and Design. Forty student volunteers from introductory psychology classes participated as *S*s and were assigned in order of appearance to one of four treatments. Two *S*s were dropped because of what appeared to be an antagonistic attitude toward the task, and were subsequently replaced. The experimental design was a 4 x 8 repeated measures factorial design, including four levels of MF (0, 10, 20, and 30 per cent), and eight successive 10-trial blocks.

Materials and Apparatus. The stimuli consisted of three-digit numbers, 39 of which ended with an even digit, and 41 of which ended with an odd digit, unsystematically distributed over the 80 trials. These stimuli were placed on projector slides, and *S*s responded by pressing buttons on a control panel labeled "X" and "Not X." In addition to the control panel, the apparatus consisted of five major components: (a) a Kodak Carousel slide projector used to present the series of stimuli (b), a 7 x 8-in. translucent screen placed directly in front of *S* on which the stimuli were rear-projected, (c) an Esterline-Angus operations recorder used to record the correct answer, *S*'s response, and the response latency, (d) three Hunter timers used to control delay and duration of feedback and post-feedback interval, and (3) a Western Union tape transmitter, used to control S's signal lamps. Delay of feedback was 0.5 sec., duration of feedback was 2.0 sec., and post-feedback interval was 2.3 sec.

During each trial the following sequence of events took place. When S pressed one of the response buttons, the response was recorded by the Esterline-Angus recorder, and the timing unit was activated. The timing unit advanced the Western Union tape transmitter, which both illuminated the correct signal lamp, and recorded the correct response, 0.5 sec. after S's response. After 2 more seconds, the timing unit again activated the transmitter, turning off the signal lamp. After another 2.3 sec., the projector advanced, and projected the next stimulus upon the screen. The Esterline-Angus recorder operated continuously throughout each S's session.

Several constraints were imposed on the sequence of stimuli in addition to approximately one half of them being even, and one half odd. The randomly selected numbers were corrected so that all digits appeared approximately equally often in all positions (first, second, and third digit spaces), excepting zero which was confined to the second and third digit spaces. The stimuli were arranged so that approximately one half of the Type X's fell in the first 40 trials, and the remainder in the second 40 trials. Runs of the same category (X or Not-X) were limited to a maximum of six successive stimuli (See Appendix 1).

Task and Procedure. The task was similar to a standard concept-formation task in that S was required to learn to classify a series of visually presented stimuli into categories labeled X and Not-X.

The problem was unlike most concept-formation problems in that the number of dimensions available to S was extremely large, even though only one was relevant. Thus it was possible for S to hypothesize the relevant dimension as being numbers above or below 500, the highest or lowest number in a certain position, some arithmetical manipulation, and so on, before finding the optimal solution, i.e., that numbers ending with an even digit (0,2,4,6,8) were Type X. The categories of X and Not-X were chosen in order to avoid S's development of two concepts, as might have occurred with categories such as X and Y.

At the beginning of the experiment, each S was given oral instructions, which were the same for all four groups. These instructions are given in Appendix 2. Two examples were given of the type of stimuli to be presented. The Ss were told to look for a rule which would allow them to solve the problem. No hint of the possibility of MF, or of the type of rule which was necessary was given. The Ss were also instructed that they would be periodically interrupted in their task to be questioned about the strategy, or decision rule which they were using at the time. The stimuli were rear-projected, one at a time, on the screen before S.

To categorize a number, S responded by pressing either the X button on the Not-X button on the control panel below the screen. The correct answer was then indicated by the illumination of a green jewel lamp above the proper button. All problems were self-paced, S being allowed as much time as needed to make any response. Each S had 80 trials, whether or not he found the optimal solution to the problem. After every five trials *S* was asked to pause and tell E what kind of a decision rule he was applying. His response was recorded by E on a scoring sheet, but no clues were given by E as to the appropriateness of *S*'s approach. At the end of 80 trials, *S* was asked for an estimate of what per cent of the stimuli had been Type X. The *S* was then told that the experiment was concluded, the answer to the problem, and the 10, 20, and 30 per cent MF groups of the existence of MF.

Results

Error Data. Here we are concerned with "true," or *absolute* scoring pertaining to the correct hypothesis, as opposed to error and correct feedback *indicated* to *S*. Absolute and indicated errors sometimes differed because of the presence of MF. Table 1 shows the summary of analysis of variance of absolute errors. Proportion of MF was significantly related to the number of errors; the more MF, the greater the proportion of absolute errors. Significance of the 10-trial Blocks effect demonstrated learning over the 80 trials. The P x B interaction, also significant, expresses the divergence of the four groups from slightly better than chance responding to their final levels. A performance level indicative of probability matching was not in evidence here (See Fig. 1). However, analysis of variance of the last block of 10 trials showed the P effect to be significant beyond the .01 level (Table 2), and the four groups were ordered in accord with their respective levels of MF.

Table 1

Analysis of Variance Summary Table for Absolute Errors in Experiment I (cue-contingent situation)

Source	df	Mean Square	F Ratio
Proportion MF (P)	3	82.64	10.57***
Ss/P	36	7.82	
10-Trial Blocks (B)	7	32.11	12.49***
P x B	21	4.13	1.61*
Ss/P x B	252	2.57	
Total	319		

Table 2

Analysis of Variance Summary Table for Absolute Errors in the Last Block of Ten Trials of Experiment I

Source	df	Mean Square	F Ratio
Proportion MF (P)	3	28.03	8.87***
Ss/P	36	3.16	
Total	39		

*Significant beyond the .05 level.

***Significant beyond the .001 level.

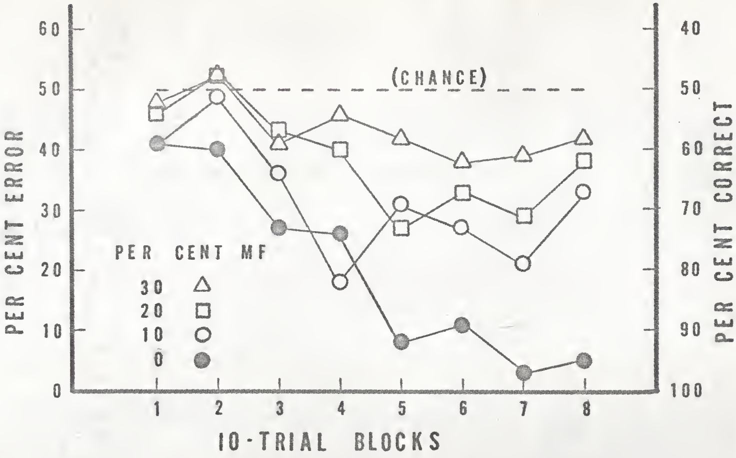

Fig. 1. Absolute error over trials by per cent MF, Expt. 1.

Latency Data. The summary of analysis of variance of latencies across 80 trials is shown in Table 3. Each S was given four scores: mean latency after an indicated error, and mean latency after an indicated correct response, for the first and the second 40 trials. These scores were used in the analysis. Proportion of MF did not significantly affect response latency. The after-error, after-correct (E) effect, on the other hand, shows that S tended to respond more slowly after an error than after a correct response (Means: 7.83 sec., and 6.51 sec.). Average latency did not change from the first 40 trials to the latter 40, but was remarkably stable, as shown by the low mean square of the Blocks effect. None of the interactions was significant. Analysis of variance of the first 40 trials only (Table 4) was done to include the zero percent MF group, which had been excluded from the first analysis since many of these Ss made no errors in the last 40 trials. Much the same results were found, the only significant affect being longer response latency following an indicate error (Means: 7.88 sec., and 6.45 sec.).

Hypothesis Data. As mentioned above, Ss were requested to verbally state their strategies after every fifth trial. This material was analyzed in terms of (a) major changes of hypotheses, (b) minor amendments of an earlier hypothesis, and (c) changes and amendments combined for total alterations of hypotheses. Tables 5, 6, and 7 summarize these analyses. In none of them was proportion MF a significant variable. The Trials effect was significant with regard to hypothesis changes and total alterations, showing a decline as Ss approached problem "solution." The first block (Trial 5) was omitted from analysis

Table 3

Analysis of Variance Summary Table for Response Latencies in Experiment I (cue-contingent situation)

Source	df	Mean Square	F Ratio
Proportion MF (P)	2	260084.7	1.04
Ss/P	27	249772.3	
After Error, or Correct (E)	1	530270.4	11.15**
Blocks of 40 Trials (B)	1	9.2	----
P x E	2	13914.6	----
P x B	2	100833.7	2.12
E x B	1	3091.4	----
P x E x B	2	27622.6	----
Residual	81	47542.3	
Total	119		

Table 4

Analysis of Variance Summary Table for Response Latencies in the First Block of Forty Trials in Experiment I

Source	df	Mean Square	F Ratio
Proportion MF (P)	3	39298.	----
Ss/P	36	86924.	
AE vs. AC (E)	1	604476.	30.86***
P x E	3	20334.	1.04
Residual	36	19586.	
Total	79		

**Significant beyond the .01 level.

***Significant beyond the .001 level.

Table 5

Analysis of Variance Summary Table for Major Hypothesis Changes by Subjects in Experiment I

Source	df	Mean Square	F Ratio
Proportion MF (P)	2	0.725	----
Ss/P	27	1.359	
Trials (T)	14	1.266	10.639***
P x T	28	.077	----
Ss x T (residual)	378	.119	
Total	449		

Table 6

Analysis of Variance Summary Table for Minor Hypothesis Amendments by Subjects in Experiment I

Source	df	Mean Square	F Ratio
Proportion MF (P)	2	0.800	1.347
Ss/P	27	0.594	
Trials (T)	14	0.260	1.405
P x T	28	0.193	1.043
Ss x T (residual)	378	0.185	
Total	449		

***Significant beyond the .001 level.

Table 7

Analysis of Variance Summary Table for Total Alterations of Hypotheses (changes and amendments combined) in Experiment I

Source	df	Mean Square	F Ratio
Proportion (P)	2	0.950	1.316
Ss/P	27	0.722	
Trials (T)	14	1.121	7.232***
P x T	28	0.165	1.065
Ss x T (residual)	378	0.155	
Total	449		

***Significant beyond the .001 level.

since at this point all Ss stated a new (first) hypothesis. The 5-trial blocks effect was not significant with respect to hypothesis amendments, the number of which remained relatively constant throughout the series. These effects are shown in Fig. 2. This graph may be read inversely, as perseveration, or retention of hypotheses, which may or may not have been optimal ones.

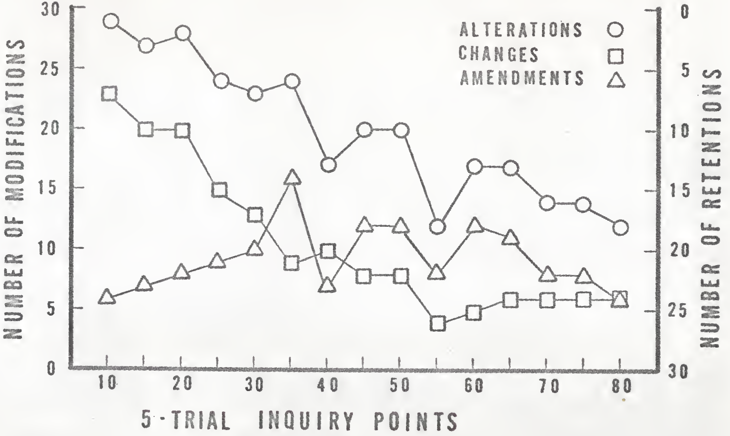

Fig. 2. Hypothesis modifications over trials, Exp. I.

CHAPTER 3

EXPERIMENT 2

NONCONTINGENT FEEDBACK

Method

Subjects and Design. The Ss were 30 student volunteers from introductory psychology classes, and as before, were assigned to groups in order of appearance. Two Ss were dropped from the experiment because of an apparently negative attitude toward the task, and were replaced. The experimental design was a 3 x 8 repeated measures factorial design. In this experiment the feedback was noncontingent with regard to the stimuli, and S's responses. No matter what S did, he was told that he was wrong a certain percentage of the time. The indicated errors occurred at the same points as did MF in Experiment 1. The three groups differed in that they consisted of 10, 20, or 30 per cent error feedback. Zero per cent error feedback was not used since it was presumed that S would soon become suspicious. There was no true or optimal answer to the problem, but any strategy S chose was as good as any other.

Materials and Apparatus. The stimuli, number of trials, and sequence of events were exactly the same as in Experiment 1. Apparatus was identical with the exception of the control panel, which had to be adapted for noncontingent feedback. The same two response buttons, labeled in the same way, were present, but in the vertical plane between them were mounted (at the top) a green lamp labeled "right," and (at the bottom) a red lamp labeled "wrong."

Task and Procedure. The task, as S understood it, was exactly the same as in Experiment 1. The S was to categorize the three-digit stimuli as X or Not-X. Instruction and procedure were also identical to those used in Experiment 1.

Results

Error Data. Collection of data based on absolute errors, as in Experiment 1, was clearly impossible in this case. Since reinforcement was distributed unsystematically throughout the 80 trials, and any strategy was as good as any other, only indicated, as opposed to absolute errors, could occur.

Latency Data. Tables 8 and 9 show results similar to those found in Experiment 1. The only significant main effect was that of E ; latencies following errors tended to be longer than those following correct responses: means for the 80 trials were 7.85 sec., and 6.17 sec., while means for the first 40 trials only were 7.75 sec., and 6.12 sec., respectively. Again, the P and B variables, as well as all interactions, were not significant.

Hypothesis Data. Analysis of variance of hypothesis changes, amendments, and total alterations (See Tables 10, 11, and 12) may be summarized in much the same way as in Experiment 1. Major changes, and total alterations decreased significantly over trials, while amendments continued to fluctuate. All other effects were not significant (See Fig. 3). However, this experimental paradigm afforded the opportunity to examine hypothesis behavior more carefully than in Experiment 1.

Table 8

Analysis of Variance Summary Table for Response Latencies in Experiment 2 (cue-noncontingent situation)

Source	df	Mean Square	F Ratio
Proportion Error (P)	2	676270.2	1.573
Ss/P	27	430012.4	
After Error, or Correct (E)	1	844201.8	14.814***
Blocks of 40 Trials (B)	1	6063.4	----
P x E	2	56492.2	----
P x B	2	146659.5	2.574
E x B	1	735.1	----
P x E x B	2	65624.4	1.152
Residual	81	56986.4	
Total	119		

Table 9

Analysis of Variance Summary Table for Response Latencies in the First Block of Forty Trials in Experiment 2

Source	df	Mean Square	F Ratio
Proportion Error (P)	2	156158.15	----
Ss/P	27	169841.97	
After Error, or Correct (E)	1	397557.60	13.25**
P x E	2	19232.55	----
Residual	27	30011.23	
Total	59		

**Significant beyond the .01 level.

***Significant beyond the .001 level.

Table 10

Analysis of Variance Summary Table for Major Hypothesis
Changes by Subjects in Experiment 2

Source	df	Mean Square	F Ratio
Proportion Error (P)	2	.260	----
Ss/P	27	.406	
Trials (T)	14	.852	11.833^{***}
P x T	28	.055	----
Ss x T (residual)	378	.072	
Total	449		

Table 11

Analysis of Variance Summary Table for Minor Hypothesis
Amendments by Subjects in Experiment 2

Source	df	Mean Square	F Ratio
Proportion Error (P)	2	.445	1.165
Ss/P	27	.382	
Trials (T)	14	.286	1.382
P x T	28	.233	1.126
Ss x T (residual)	378	.207	
Total	449		

***Significant beyond the .001 level.

Table 12

Analysis of Variance Summary Table for Total Alterations of Hypotheses (changes and amendments combined) in Experiment 2

Source	df	Mean Square	F Ratio
Proportion Error (P)	2	1.285	2.156
Ss/P	27	.596	
Trials (T)	14	.481	2.237**
P x T	28	.196	----
Ss x T (residual)	378	.215	
Total	449		

**Significant beyond the .01 level.

1. Failure to conform to verbalized hypotheses. In the groups with 10 and 20 per cent error feedback there was a total of eleven blocks of five errorless trials, the same for all Ss within groups. Thus there were 550 points to check whether S, in the absence of an error, always followed his verbalized hypothesis. In fact, in 105 cases (19.09 per cent) Ss did not conform to their hypotheses. This, qualitatively, is not a unique finding (e.g., Suppes and Schlag-Rey, 1965), although its quantification may be of value. However, such a measurement may also be considered an indicant of how well Ss verbalized their hypotheses.

2. Spontaneous hypothesis alteration. The same eleven blocks of five errorless trials allowed an examination of hypothesis changes and amendments not precipitated by an error. With ten Ss in each group, there were 110 opportunities to test the assumption that

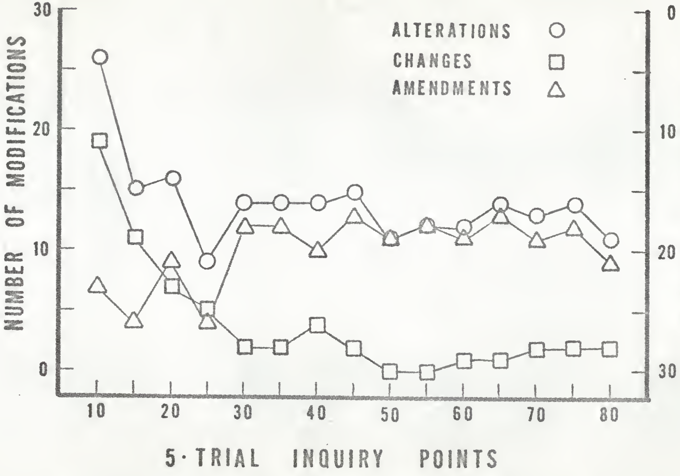

Fig. 3. Hypothesis modifications over trials, Exp. 2.

no such changes will occur. Upon inspection the data revealed eleven major changes of hypotheses (10 per cent), and 37 amendments to the prior hypothesis (33.6 per cent), for a total of 48 alterations of hypotheses (43.6 per cent), without errors in the preceding block of five trials.

3. *Hypothesis perseveration*. In the noncontingent data there were 34 blockscontaining at least one indicated error. It follows that there was a total of 340 inquiry points available to check hypothesis perseveration following an error. In 178 cases (52.35 per cent), *S* did not change or amend his hypothesis, but maintained the decision rule he had verbalized during the previous inquiry.

CHAPTER 4

EXPERIMENT 3

OCCURRENCE PREDICTION

Method

Subjects and Design. Again, 30 introductory psychology students were assigned to groups in order of appearance, and the experimental design was a 3 x 8 repeated measures factorial design. In this experiment the program tapes from Experiment 2 were used with the control panel of Experiment 1 in such a manner that Type Not-X was the correct answer on 10, 20, or 30 per cent of the 80 trials, and occurred at the same points as did MF in Experiment 1, and noncontingent errors in Experiment 2.

Materials and Apparatus. The apparatus and sequence of events were the same as in Experiment 1, with one exception. Instead of the presentation of three-digit stimuli, the screen was blacked out, and a red jewel lamp placed in its center. The only function of the lamp was to tell S when he could respond.

Task and Procedure. In contrast to the plethora of dimensions available to S for his categorization rules in Experiments 1 and 2, here there were essentially none. On each trial S was required to push a button to predict the occurrence of either an X or a Not-X event. All Ss were provided with the same instructions (See Appendix 2), which explained that their task was to learn to predict whether each trial (event) would be an X or a Not-X, and that the correct answer would be shown on the control panel after their response. Again, they were told that there would be a periodic inquiry regard their hypotheses. Absolutely no hints were given as to the type of hypothesis they were to use, or that the series was unsystematic. All other procedures were identical to those of Experiments 1 and 2.

Results

Error Data. The summary of analysis of variance of absolute errors is presented in Table 13. Here an absolute error is a Not-X response. Proportion of Not-X was seen to be a significant source of variance, inversely related to the proportion of correct responses. These date are graphed in Fig. 4. Improvement of performance under all Not-X proportions was reflected by the significant Blocks effect. The significant P x B interaction showed that the divergence of the three curves over 80 trials wes also significant. The final proportions of correct responses approximated probability matching behavior.

Latency Data. As might be expected from the results of Experiments 1 and 2, Proportion had no effect (See Teble 14). Response latency was significantly longer following an indicated error than an indicated correct response (Means: 4.44 sec., and 3.76 sec.). The Blocks effect was also significant, with a slightly longer average latency in the second 40 trials (Means: 3.96 sec., and 4.25 sec.). The significant P x E interaction (See Fig. 5) was an unusual one. It would appear that after a correct response, latency dropped in direct proportion to the amount of Not-X events. Latencies after errors, however, were lowest at 20 per cent and almost equally high

Table 13

Analysis of Variance Summary Table for Absolute Errors in Experiment 3 (cueless situation)

Source	df	Mean Square	F Ratio
Proportion Not-X (P)	2	108.71	17.94***
Ss/P	27	6.06	
Ten-Trial Blocks (B)	7	26.86	13.92***
P x B	14	5.57	2.886**
Ss/P x B	189	1.93	
Total	239		

Table 14

Analysis of Variance Summary Table for Response Latencies in Experiment 3 (cueless situation)

Source	df	Mean Square	F Ratio
Proportion Not-X (P)	2	34458.3	1.324
Ss/P	27	26032.1	
After Error or Correct (E)	1	137905.2	24.856***
Blocks of 40 Trials (B)	1	25872.0	4.663*
P x E	2	30048.4	5.416**
P x B	2	5270.8	----
E x B	1	36192.2	6.523*
P x E x B	2	5397.8	----
Residual	81	5548.1	
Total	119		

*Significant beyond the .05 level.

**Significant beyond the .01 lvel.

***Significant beyond the .001 level.

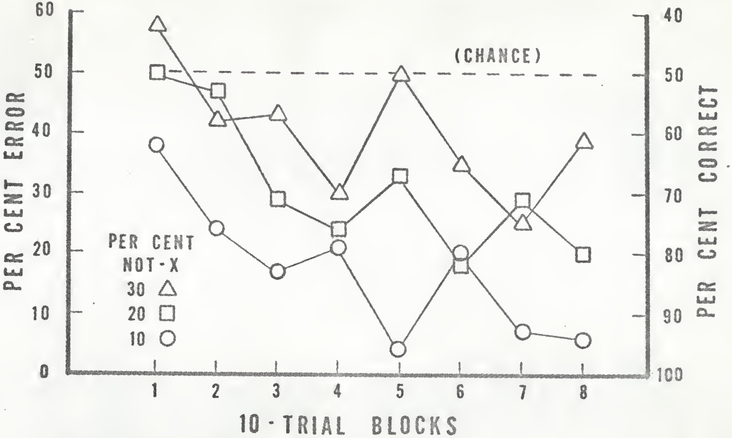

Fig. 4. Absolute error over trials by per cent Not-X event , Exp. 3.

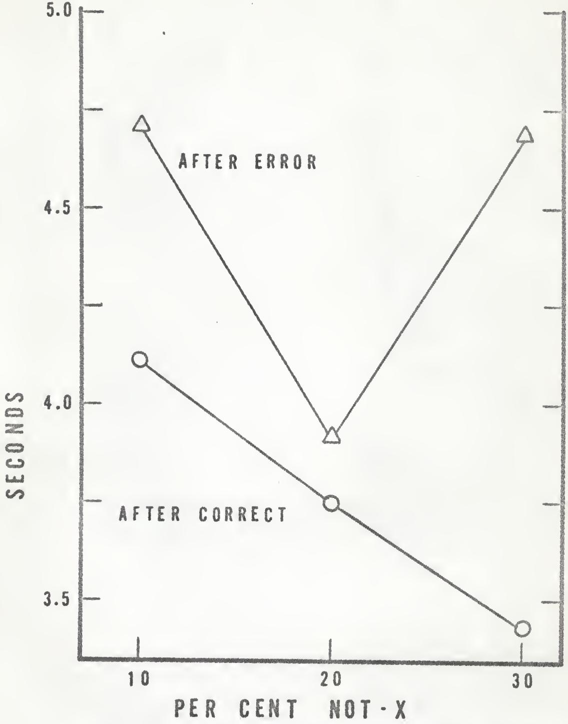

Fig. 5. The interaction of per cent Not-X events and the after-error, after-correct latency effect, Exp. 3.

at 30 and 10 per cent Not-X event proportions. Such results tend to suggest the operation of more than one factor controlling response latency. The E x B interaction (Fig. 6) indicated that latency after correct responses dropped over trials, while latency after an indicated error increased as the problem went on. Analysis of the first 40 trials only (Table 15) revealed similar effects, with E and P x E significant for the same reasons.

Hypothesis Data. Since Ss under these conditions changed their strategy at almost every inquiry point, hypotheses could not be analyzed by the same methods as in Experiments 1 and 2. All Ss, without exception, attempted to use sequential rules based on the series of events observed by them prior to inquiry.

Table 15

Analysis of Variance Summary Table for Response Latencies in The First Block of Forty Trials in Experiment 3

Source	df	Mean Square	F Ratio
Proportion Not-X (P)	2	7380.2	----
Ss/P	27	9587.5	
After Error or Correct (E)	1	16401.0	12.50**
P x E	2	5123.5	3.904*
Residual	27	1312.4	
Total	59		

*Significant beyond the .05 level.

**Significant beyond the .01 level.

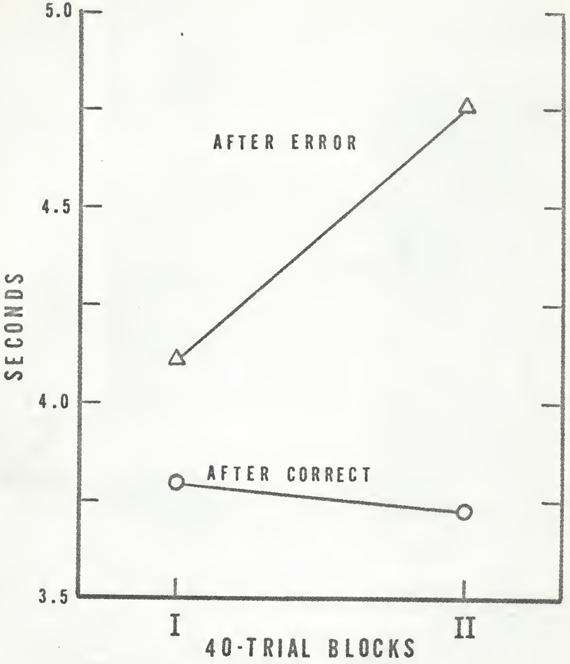

Fig. 6. The interaction of 40-trial blocks and the after-error, after-correct latency effect, Exp. 3.

CHAPTER 5

EXPERIMENTS 1, 2, AND 3

COMBINED ANALYSES

Validity

Experiments 1 and 2 were run concurrently, with Ss assigned alternately to the two experiments. Thus it is reasonable and entirely valid to perform combined analyses on these experiments. Experiment 3 was run later, and hence the interpretation of combined analyses must be tempered by the consideration that there may have been differences in the S population. Such differences, if indeed any exist, are considered minor.

Results

Error Data. It is appropriate that absolute error rates be compared between Experiments 1 and 3. Table 16 shows that with the two experiments combined in a single analysis of variance, Groups (cue vs non-cue), 10-trial Blocks, and the G x B interaction were all significant sources of variance. These effects are shown in Fig. 7. After the first block of 10 trials the cue group (Experiment 1) remained worse than the cueless group (Experiment 3) in number of absolute errors. Both groups' performances showed improvement, but the cueless group improved at a greater rate. Proportion (MF or Not-X) also had a significant effect, being related to total number of absolute errors (Means: 33.6, 28.0, and 19.7). The three-way interaction of G x P x B is somewhat difficult to interpret, but

Table 16

Analysis of Variance Summary Table for Absolute Errors Based on Experiments 1 and 3 Combined (cue and cueless)

Source	df	Mean Square	F Ratio
Groups I and III (G)	1	87.55	12.168**
Proportion MF and Not-X (P)	2	122.28	16.995***
G x P	2	13.64	1.896
_S_s/G x P	54	7.195	
Ten-Trial Blocks (B)	7	37.23	15.951***
G x B	7	5.51	2.361*
P x B	14	2.68	1.148
G x P x B	14	5.31	2.275**
Residual	378	2.334	
Total	479		

*Significant beyond the .05 level.

**Significant beyond the .01 level.

***Significant beyond the .001 level.

appears to indicate an increasing divergence of the six curves over the 80 trials beyond that explainable by the P effect alone.

Latency Data. Response latencies were the only behavioral measure applicable to comparison of all three experimental groups. Table 17 presents the summary of analysis of variance for their combined data. Once again, Proportion (MF, NC errors, Not-X events) did not prove to have a significant effect upon response latency. The significant Groups effect seemed to result mainly from the rapid responses of _S_s in

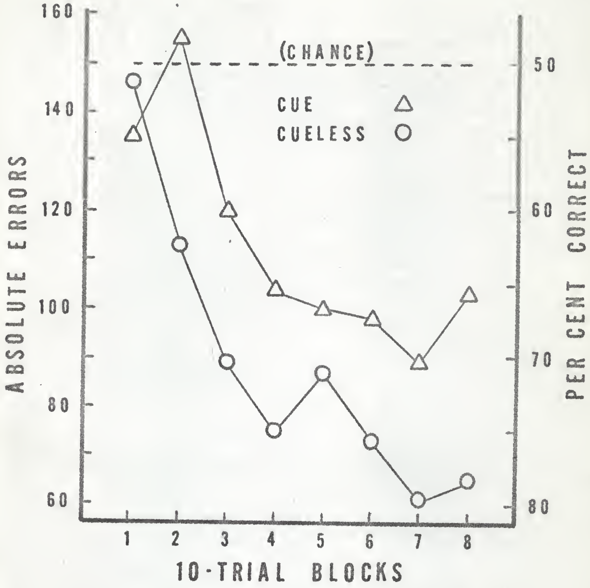

Fig. 7. The effects of groups, trials, and their interaction on absolute error, Exps. 1 and 3 combined.

Table 17

Analysis of Variance Summary Table for Response Latencies in Experiments 1, 2, 3 Combined

Source	df	Mean Square	F Ratio
Proportion (P) (MF, Not-X, Error)	2	190840.	----
Groups I, II, III (G)	2	5515315.	23.440***
P x G	4	389365.	1.654
<u>S</u>s/ P x G	81	235287.	
After Error or Correct (E)	1	1353014.	36.872***
Block of 40 Trials (B)	1	20071.	----
P x E	2	103.	----
P x B	2	145237.	3.958*
E x B	1	9120.	----
G x E	2	78332.	2.134
G x B	2	6530.	----
P x E x B	2	30150.	----
G x P x E	4	50285.	1.370
G x P x B	4	53959.	1.470
G x E x B	2	16150.	----
G x P x E x B	4	34261.	----
Residual	243	36694.	
Total	359		

*Significant beyond the .05 level.

***Significant beyond the .001 level.

Experiment 3. The mean latencies were: Expt. 1, 7.17 sec., Expt. 2, 7.01 sec., Expt. 3, 4.10 sec. As in the previous analyses, the E effect was a highly significant source of variance, response latency after an indicated error being longer than after an indicated correct button-press (Means: 6.47 sec., and 5.24 sec.).

A most interesting effect was shown by the significant P X B interaction (Fig. 8). It would appear that over trials the latency of any response tended to grow longer in the 30 per cent groups, remain about the same in the 20 per cent groups, and decrease in the 10 per cent groups. Although this effect did not approach significance in the single experiment analyses, the general trend was in this direction.

An analysis of variance for <u>total</u> (as opposed to mean) latencies over all three experiments, to further investigate G, P, and G x P effects yielded only Groups as a significant source of variance. As a final check, two more separate analyses were run on (a) mean latency following indicated errors only, and (b) mean latency following indicated correct responses only. It was reasoned that in this way, a Proportion effect otherwise hidden by unequal numbers of errors in the different proportions might appear. Once again, however, only the Groups effect proved to be significant.

<u>Hypothesis Data</u>. These measurements could be compared only between the Experiments 1 and 2. Summaries of analyses of variance are presented in Tables 18, 19, and 20 for major hypothesis changes, hypothesis amendments, and total alterations of hypotheses, respectively. None of the analyses revealed a significant source of variance in Proportion.

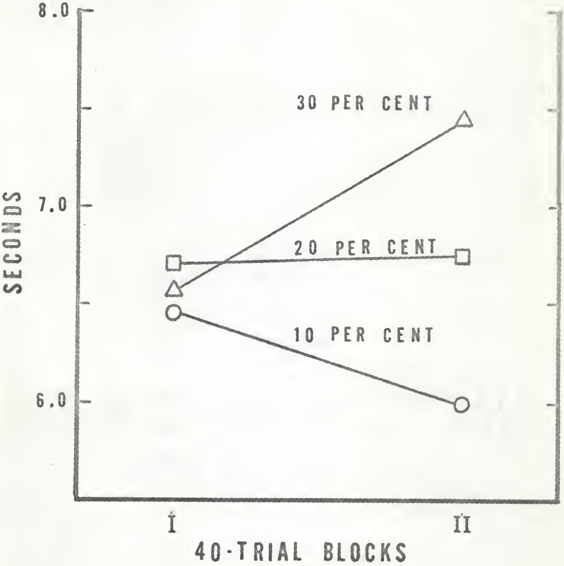

Fig. 8. The interaction effect of 40-trial blocks and per cent MF, NC errors, and Not-X events upon latency.

Table 18

Analysis of Variance Summary Table for Major Hypothesis Changes by Subjects in Experiments 1 and 2 Combined

Source	df	Mean Square	F Ratio
Proportion MF or Error (P)	2	.335	----
Groups I and II (G)	1	10.890	12.347***
P x G	2	.650	----
Ss/ P x G	54	.882	
Trials (T)	14	1.971	20.747***
P x T	28	.068	----
G x T	14	.147	1.547
P x G x T	28	.064	----
Ss x T (residual)	756	.095	
Total	899		

***Significant beyond the .001 level.

On the other hand, the Trials effect was significant in all analyses, the changes and total alterations decreasing, while amendments increased over the eight 10-trial blocks (See Fig. 9). The Groups effect was significant in the change and alteration analyses, the noncontingent group making consistently fewer changes and total alterations, while the groups did not differ in amendment rates. The G x T interaction was significant in the total alteration analysis, showing an increasing, then decreasing disparity in total number of alterations of hypotheses across the series of 80 trials.

Table 19

Analysis of Variance Summary Table for Minor Hypothesis Amendments by Subjects in Experiments 1 and 2 Combined

Source	df	Mean Square	F Ratio
Proportion MF or Error (P)	2	1.140	2.336
Groups I and II (G)	1	.120	----
P x G	2	.105	----
Ss/ P x G	54	.448	
Trials (T)	14	.409	2.087**
P x T	28	.188	----
G x T	14	.136	----
P x G x T	28	.238	1.214
Ss x T (residual)	756	.196	
Total	899		

**Significant beyond the .01 level.

Table 20

Analysis of Variance Summary Table for Total Alterations of Hypotheses (changes and amendments combined) in Experiments 1 and 2

Source	df	Mean Square	F Ratio
Proportion MF or Error (P)	2	1.470	2.230
Groups I and II (G)	1	8.800	13.354***
P x G	2	.765	1.161
Ss/ P x G	54	.659	
Trials (T)	14	1.189	6.427***
P x T	28	.171	----
G x T	14	.413	2.232**
P x G x T	28	.191	1.032
Ss x T (residual)	756	.185	
Total	899		

**Significant beyond the .01 level.

***Significant beyond the .001 level.

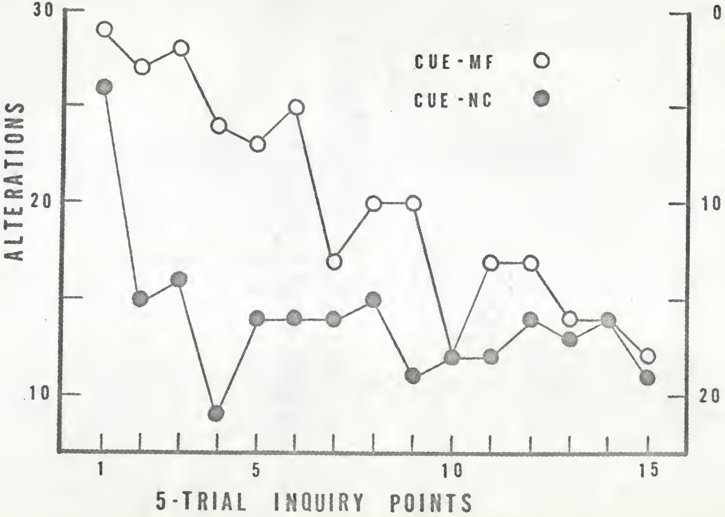

Fig. 9. The effects of groups, trials, and their interaction upon hypothesis modifications, Exps. 1 and 2 combined.

CHAPTER 6

DISCUSSION

Probability Matching

The absolute error results indicated that in both Experiments 1 and 3, Ss showed learning during the series of trials. Proportions of MF and Not-X events governed the rate and the final extent of this learning. Probability matching, however, occurred only in the occurrence prediction situation, while only something akin to it appeared in the multi-dimensional MF experiment. It is possible that with more trials, Ss in Experiment 1 might have reached probability matching, and that with still more, they might have adopted a maximizing solution, as did Bourne's (1963) Ss.

Since the hope of a unique solution was present in both experiments, it may be that such an expectancy is a necessary, but not sufficient prerequisite of probability matching behavior. It is clear from inspection of the data that the probability matching found in Experiment 3 was caused by neither a "win-stay, lose-shift hypothesis, nor by a methodological artifact. On the other hand, it may be explained by the fact that all Ss made use of sequential rules based upon the series of events prior to each inquiry trial. Indeed, Ss could have used almost no other kind of rule. Since prior events almost invariably fall into proportions of roughly 10, 20, and 30 per cent Not-X events, it is not surprising that Ss, having adopted a sequential hypothesis based on the sequence of prior events, would evidence probability matching behavior. The simplest case, for example, is the one in which S responds as if he were repeating the previous five or ten trials.

It must be recalled, however, that the experimental procedure required S to state a hypothesis every five trials, forcing him into the role of a hypothesis tester, which otherwise he might not have assumed.

To test the assertion that Ss are aware of the overall frequency of the two events, Ss were asked at the end of the experiment what per cent of the stimuli had been Type X. In Experiment 3, the cueless situation, Ss' estimates averaged 72.1, 81.5, and 90.4 per cent, which were quite close to actuality. In Experiment 1 the mean estimate was 58.75 per cent, somewhat above the true figure of 48.75 per cent Type X. In both experiments, however, Ss needed a considerable amount of time to make this apparently difficult estimation, demonstrating that such estimates were not used as general guides, at least not at a fully conscious (deliberate) level of awareness. Perhaps the term "probability learning" is then misleading, and should be replaced by "learning with probabilistic cues," as has been suggested by Bruner et al. (1956).

Besides probability matching, there was another difference between the results of Experiments 1 and 3. The Ss in the occurrence-predicting, cueless situation learned at a significantly faster rate than did Ss who had to deal with MF and numerous irrelevant dimensions. It is quite possible that these two differences, learn-rate rate, and presence of probability matching, are related. Exactly how many dimensions were sampled by Ss in Experiment 1 cannot be estimated with any degree of precision. Nevertheless, there was an average of 5.3 major hypothesis changes per S. Such a figure, of course, does not include dimensions sampled and discarded between inquiry points, or those observed, but not directly tested. It is more than likely that this multiplicity of irrelevant dimensions caused the decrement in rate of learning (See for example Pishkin, 1960: Johannsen, 1962), as well as the failure to demonstrate probability matching in Experiment 1. The Ss in Experiment 1, seeking a complete solution to the problem, continued to attend to (or sample) worthless cues, slowing their learning, while Ss in Experiment 3 had nothing to distract them from the simplest kind of sequential strategies. Thus one is led to believe that probability matching is apt to take place in occurrence predicting, or other simple situations in which rules cannot reach a high degree of complexity. The data from this study suggest that probability matching in occurrence prediction results from the use of sequential hypotheses.

Behavioral Characteristics of Problem Solving

In all three experiments, mean response latency after an indicated error was significantly longer than after an indicated correct response. These results are in substantial agreement with those of Erickson et al. (1965), who measured latencies in a standard concept formation task. However, mean latencies of Experiments 1 and 2 of the present study were considerably longer (about 3 sec.) than those of Erickson's experiments, while those of Experiment 3 were approximately the same.

Also, Erickson found a decreasing latency over trials which occurred in this study only in the 10 per cent groups for the combined experiments. These discrepancies are probably the result of the present use of probabilistic feedback and the large number of possible dimensions as opposed to the standard task with only 4 dimensions used by Erickson. On the other hand, it is possible that differences in method and apparatus account for these discrepancies. At any rate, the differences in mean latencies after indicated errors and indicated correct responses tend to support the hypothesis-resampling assumptions of the mathematical models.

One may ask, however, whether hypothesis resampling is the only possible explanation of longer latencies after errors. It is conceivable that an aversive effect of being told one is "wrong" might account for the response-latency results. Byers (1965) has done work suggesting that E's invalidations of S's hypotheses have the effect of cumulative punishment. If this is the case, longer latencies after errors might well be expected. Such a viewpoint would also help to explain the E x B interaction effect in Experiment 3 (Fig. 6), in which mean latency after an indicated error increased with the number of trials, and the P x B effect in the combined analysis (Fig. 8), in which mean latencies increased or decreased over trials as a function of per cent MF, NC error, and Not-X events. If hypothesis resampling were the only reason for latency effects, there would be no parsimonious explanation for these Blocks interaction effects. If, on the other hand, being told one is "wrong" has a cumulative aversive effect, increasing latency with a mounting number of errors over the 80 trials could be predicted.

The P x E interaction in Experiment 3 (Fig. 5) may indicate that both effects, resampling and aversive, were operating in the cueless situation, governing the high 10 and 30 per cent Not-X after-error latencies, while the 20 per cent group was not so strongly influenced by either of them. In other words, it may be speculated that while latency increases after errors, this increase is a result of two effects: (a) a positively accelerated increase of aversive effect from 10 to 30 per cent Not-X and (b) a negatively accelerated increase of the resampling effect for 10 to 30 per cent Not-X.

Inspecting the rates of total hypothesis alterations in Experiments 1 and 2, it is seen that the G x B interaction expresses similar rates at the beginning, followed by a sharper drop of the noncontingent group. The alteration rates of this group remained below those of Ss in Experiment 1 until the last few trials, at which time the curves again converged (Fig. 9). This effect would be most simply explained by calling upon the idea of hypothesis modification as an indirect function of errors. That is, hypothesis alterations may not be made immediately following an error, but increase with the overall amount of error feedback. Thus the rapid drop of the noncontingent group Ss is explained by the fact that they have already reached their maximum possible performance, while the cue-MF group reached this level later in the series, when the curves again converged. However, it may be argued that if hypothesis change is a function of some level of indicated error feedback, Proportion should be a significant source of variance in all the hypothesis data analyses. Higher P causes more errors, and thus should produce more hypothesis changes and amendments. Nonsignificance of the P effect in these analyses seems to belie the explanation of hypothesis modification as a result of increasing errors.

This argument may possibly be countered with the supposition that the P effect was not significant because the difference in the number of errors between the three proportions was too slight (10 per cent in the noncontingent group, and less than that in the MF group until finding the optimal solution). But the difference of average number of indicated errors between groups at the beginning was about 30 per cent (Exp. 1, 50 per cent; Exp. 2, 20 per cent). The average indicated error difference between groups generally remained larger than that between proportions until the last few 10-trial blocks. The Ss in Experiment 1, until they discovered the optimal solution could do no better than 50 per cent correct, and had to continue to alter their hypotheses if they were to do better than chance. The Ss in Experiment 2, however, soon found that their hypotheses were yielding what appeared to be a partial solution, and thus could continue to amend, but seldom completely changed their hypotheses. One thus tends to suspect that hypothesis alterations occur not necessarily immediately following an error, but only when the level of error is uncomfortably high. This unique effect cannot generally be observed in a standard concept formation task, wherein S usually rises from chance performance to complete solution, without the intervening stage provided by probabilistic feedback.

It is rather interesting to note that Ss in the noncontingent group rapidly adopted firm hypotheses, and changed them less than did Ss in Experiment 1, who had to deal with a problem in which there existed a single optimal hypothesis. As mentioned before, in Experiment 2 any hypothesis was as good as any other, and the almost immediate "freezing," or adoption of a single hypothesis (allowing "exceptions to the rule"), fits the definition of superstitious behavior; continued use of a hypothesis which actually works no better than chance, or guessing behavior.

It is evident from the results of all three experiments, and especially those of Experiment 2, that Ss do retain their hypotheses although they are making a certain number of errors. Any hypothesis that leads to better than chance performance is superior to no hypothesis at all, and is not dispensed with because some errors are occurring. Instead, Ss often attribute errors to some unknown exception to the rule.

Probability learning could not take place without some perseveration in the face of error. If hypotheses changed with every error, it would not be possible to learn and maintain a response unless the situation provided at least one "perfect" answer, which is seldom the case outside of the laboratory. Restle (1962) is cognizant of this fact, and is able to incorporate perseveration into his mathematical learning model.

On the other hand, neither Restle's nor Atkinson, Bower, and Crother's models assume, or account for hypothesis changes when no error has occurred. Data from Experiment 2 show that hypothesis change without immediately preceding error is, in fact, a not uncommon occurrence. It would appear then, that although the mathematical model approach to problem-solving is able to describe final error and trial data closely, it makes unwarranted assumptions regarding the behaviors through which these data are produced.

While it has been possible to make some informative comparisons among these three models of learning, it is noteworthy in itself that the general behaviors and results do not greatly differ between these three experimental paradigms. There are, of course, clear differences in the results, and one model may be more applicable than another for investigation of certain variables because of its particular characteristics. It appears, though, that each model presents a valid approach to the study of learning with probabilistic cues.

CHAPTER 7

SUMMARY

Previous research in learning with probabilistic cues has employed three primary experimental paradigms: occurrence prediction, misinformative feedback, and noncontingent feedback. Occurrence prediction experiments require Ss to predict the occurrence of one of two stimuli over a series of trials, without physical cues to aid their predictions. Misinformative feedback (MF) is also usually a two-choice situation, but cues are presented to aid categorizations. These cues, however, are valid only some percentage of the time. Noncontingent (NC) feedback experiments also have cues, none of which are valid. Reinforcement is predetermined, unsystematic, and occurs on a percentage basis.

A general finding in studies using the first two experimental situations is that Ss respond positively in a proportion approximating the percentage of positively reinforced trials. One purpose of the present research was to investigate possible origins and generality of this probability matching behavior by examining not only Ss' error data, but also their verbalized hypotheses.

Mathematical models attempting to describe data acquired in concept-formation experiments make two basic assumptions: when S receives an error he will always change his hypothesis, but when he receives no errors he will never change his hypothesis. A second major purpose of this study was to examine error records, hypothesis statements, and response latency times of Ss to test the validity of these assumptions.

The present study consisted of three experiments; one of each of the three primary paradigms. One hundred male and female introductory psychology students at Kansas State University served as Ss. Each experiment was a repeated measures factorial design, with eight successive 10-trial blocks, and levels ranging from zero to thirty per cent MF, NC error feedback, and Not-X events. The stimuli in Experiments 1 and 2 were three-digit numbers rear-projected on a screen before S. In Experiment 3 Ss responded after the illumination of a signal lamp. The Ss indicated their categorizations by pressing one of two buttons on a control panel. Feedback was presented by jewel lamps on the same panel. After every five trials Ss stated their hypotheses regarding the problem's solution.

The error results indicated that probability matching did not occur in the MF experiment, but took place in the simple occurrence predicting situation as a result of Ss adopting sequential categorization rules. Another difference between these two experiments was that Ss in the occurrence predicting groups learned at a significantly faster rate. It may be that both differences were due to the large number of irrelevant dimensions in the MF experiment.

In all three experiments mean response latency was longer after an error than after a correct response. Reasoning that hypothesis change (resampling) consumes a certain amount of time, such results tend to support the assumptions of the mathematical models. On the other hand, it is probable that an aversive effect of being told one is "wrong" accounts for at least some of these latency data.

Hypothesis data from the NC experiment make it clear that Ss do not necessarily change their hypotheses immediately after an error. Neither do they always maintain a hypothesis in the absence of an error. Instead, as suggested by the difference in hypothesis alteration rates of the MF and NC experiments, Ss appear to modify their hypotheses only when the overall level of error is uncomfortably high. Although there were many significant differences, the general trend of behaviors and results was similar in all three of the experimental paradigms.

REFERENCES

Atkinson, R. C., Bower, G. H., and Crothers, E. J. An introduction to mathematical learning theory. New York: John Wiley & Sons, Inc., 1965.

Bourne, L. E. Long-term effects of misinformative feedback upon concept formation. J. exp. Psychol. 1963, 65, 139-147.

Bourne, L. E. Hypotheses and hypothesis shifts in classification learning. J. gen. Psychol., 1965, 72, 251-261.

Bruner, J. S., Goodnow, J. J., and Austin, R. A study of thinking. New York: John Wiley & Sons, Inc., 1956.

Byers, J. L. Hypothesis behavior in concept attainment. J. educ. Psychol., 1965, 56, 337-342.

Erickson, J. R., Zajkowski, M. M., and Ehmann, E. D. All-on-none assumptions in concept identification: analysis of latency data. Paper read at Midwest Psychol. Ass., Chicago, April, 1965.

Estes, W. K. Probability learning. In A. W. Melton (Ed.), Categories of human learning. New York: Academic Press, 1964. pp. 89-128.

Estes, W. K., and Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. J. exp. Psychol., 1954, 47, 225-234.

Friedman, M. P., Burke, C. J., Cole, M., Keller, L., Millward, R. B., and Estes, W. K. Two-choice behavior under extended training with shifting probabilities of reinforcement. In R. C. Atkinson (Ed.), Studies in mathematical psychology. Stanford: Stanford Univer. Press, 1964, pp. 250-316.

Goodnow, J. J. Response sequences in a pair of two-choice probability situations. Amer. J. Psychol., 1955, 68, 624-630. (a)

Goodnow, J. J. Determinants of choice distributions in two-choice situations. Amer. J. Psychol., 1955, 68, 106-116. (b)

Goodnow, J. J., and Pettigrew, T. F. Effect of prior patterns of experience on strategies and learning sets. J. exp. Psychol., 1955, 49, 381-389.

Goodnow, J. J., and Postman, L. Probability learning in a problem-solving situation. J. exp. Psychol., 1955, 49, 16-22.

Grant, D. A., Hake, H. W., and Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. J. exp. Psychol., 1951, 42, 1-5.

Greeno, J. G. Effects of nonreinforced trials in two-choice learning with noncontingent reinforcement. J. exp. Psychol., 1962, 64, 373-379.

Guthrie, E. R., and Horton, G. P. Cats in a puzzle box. New York: Rinehart & Co., 1946.

Hake, H. W., and Hyman, R. Perception of the statistical structure of a random series of binary symbols. J. exp. Psychol., 1953, 45, 64-74.

Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. J. exp. Psychol., 1939, 25, 294-301.

Hyman, R., and Jenkin, N. S. Involvement and set as determinants of behavioral stereotypy. Psychol. Repts., 1956, 2, 131-146.

Jenkins, H. M. An exploratory study of the development of concepts with noncontingent reinforcement. Bell Lab. tech. Memo., 1959, No. 59-122-3.

Jenkins, H. M., and Ward, W. C. Judgement of contingency between responses and outcomes. Psychol. Monogr., 1965, 79, No. 1 (Whole No. 594).

Johannsen, W. J. Concept identification under misinformative and subsequent informative conditions. J. exp. Psychol., 1962, 64, 631-635.

Levine, M. Hypothesis behavior by humans during discrimination learning. J. exp. Psychol., 1966, 71, 331-338.

Messick, S. J., and Solley, C. M. Probability learning in children: some exploratory studies. J. gen. Psychol., 1957, 90, 23-32.

Morin, R. E. Factors influencing rate and extent of learning in the presence of misinformative feedback. J. exp. Psychol., 1955, 49, 343-351.

Pishkin, V. Effects of Probability of misinformation and number or irrelevant dimensions upon concept identification. J. exp. Psychol., 1960, 59, 371-377.

Restle, F. Statistical methods for a theory of cue learning. *Psychometrika*, 1961, *26*, 291-306.

Restle, F. The selection of strategies in cue learning. *Psychol. Rev.*, 1962, *69*, 329-343.

Skinner, B. F. "Superstition" in the pigeon. *J. exp. Psychol.*, 1948, *38*, 168-172.

Suppes, P., and Schlag-Rey, M. Observable changes of hypotheses under positive reinforcement. *Science*, 1965, *148*, 661-662.

Wright, J. C. Consistency and complexity of response sequences as a function of schedules of noncontingent reward. *J. exp. Psychol.*, 1962, *63*, 601-609.

APPENDIX

APPENDIX 1

LIST OF SLIDES PRESENTED IN EXPERIMENTS 1 AND 2 AND OCCURRENCES OF THE THREE PROPORTIONS OF MF, NC ERRORS, AND NOT-X EVENTS IN EXPERIMENTS 1, 2, AND 3

1. 231
2. 554 **
3. 978
4. 389
5. 973 ***
6. 603
7. 463
8. 938
9. 495 **
10. 367
11. 709
12. 731 ***
13. 615
14. 313
15. 570 **
16. 824
17. 979
18. 967
19. 726 *
20. 610

21. 148
22. 891
23. 259 ***
24. 814
25. 165 *
26. 548
27. 580
28. 117
29. 749
30. 697 *
31. 693
32. 288
33. 204 **
34. 131
35. 726
36. 523
37. 108 ***
38. 696 *
39. 837
40. 865

41. 462 **
42. 350
43. 595
44. 182
45. 744 ***
46. 854
47. 971
48. 882
49. 317
50. 549 **
51. 183 *
52. 105
53. 324
54. 940
55. 997 **
56. 483
57. 110 ***
58. 382
59. 842
60. 379 *

61. 246
62. 448
63. 617
64. 102 *
65. 276
66. 687 ***
67. 408
68. 437 *
69. 922
70. 451
71. 330
72. 223
73. 628 **
74. 504
75. 423 ***
76. 256
77. 435
78. 362
79. 758
80. 271

*MF, Noncontingent error, or Not-X event with a proportion of 30%

** with proportions of 30% and 20%

*** with proportions of 30%, 20%, and 10%

APPENDIX 2

INSTRUCTIONS FOR EXPERIMENTS 1 AND 2

This is an experiment in numerical concept formation. Projected on the screen before you, you will see a series of three-digit numbers, such as these (Two examples given). Your task is to decide whether or not these numbers are of a certain type, which we call "Type X". In other words, you must learn what attributes the Type X's have in common.

Every time one of these numbers is presented to you, you will indicate by pressing one of the two buttons, whether the number on the screen is, or is not, Type X. The signal lights on the switch box will then tell you if your decision was right or wrong so that you will know what the correct answer was. Another number will then be presented.

INSTRUCTIONS FOR EXPERIMENT 3

This is an experiment in concept formation. You are about to see a series of events, some of which we call "Type X," and others "Type Not-X." Here's how it works: when the red light comes on, that's the signal that we're ready for the next event. Anytime after the red light appears, you are to press one of the two buttons on the control box before you to indicate your prediction.

Immediately following your response, one of the two green lights will come on to tell you what the correct answer was, so that you will know if your prediction was right. After a few seconds the red light will come on again, signaling the next event. The red light does not give any clues about the answer you are to give -- it is only a signal. Initially you will just be guessing, but as the problem goes on, you will be able to figure out better whether the next event will be a Type X or a Type Not-X.

BOTH SETS OF INSTRUCTIONS FOLLOWED BY THIS STATEMENT

Every so often I will ask you to tell me what your hypothesis is, that is, your decision rule on strategy for making these choices. Start with the simplest possible strategies and build from there. Try to get as many as possible correct. Finding the complete solution to the problem is not simple, but many students are able to do it.

Do you have any questions before we begin?

A COMPARISON OF THREE MODELS OF LEARNING WITH PROBABILISTIC CUES

by

STEVEN PATRIC ROGERS

A.B., Stanford University, 1964

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Psychology

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1966

Previous research in learning with probabilistic cues has employed three primary experimental paradigms: occurrence prediction, misinformative feedback, and noncontingent feedback. Occurrence prediction experiments require Ss to predict the occurrence of one of two stimuli over a series of trials, without physical cues to aid their predictions. Misinformative feedback (MF) is also usually a two-choice situation, but cues are presented to aid catagorizations. These cues, however, are valid only some per cent of the time. Noncontingent feedback (NC) experiments also have cues, none of which are valid. Reinforcement is predetermined, unsystematic, and occurs on a percentage basis.

A general finding in studies using the first two experimental situations is that Ss respond positively in a proportion approximating the percentage of positively reinforced trials. One purpose of the present research was to investigate possible origins and generality of this probability matching behavior by examining not only Ss' error data, but also their verbalized hypotheses.

Mathematical models attempting to describe data acquired in concept-formation experiments make two basic assumptions: when S receives an error he will always change his hypothesis, but when he receives no errors he will never change his hypothesis. A second major purpose of this study was to examine error records, hypothesis statements, and response latency times of Ss to test the validity of these assumptions.

The present study consisted of three experiments; one of each of the three primary paradigms. One hundred male and female introductory psychology students at Kansas State University served as Ss.

Each experiment was a repeated measures factorial design, with eight successive 10-trial blocks, and levels ranging from zero to thirty per cent MF, NC error feedback, and Not-X events. The stimuli in Experiments 1 and 2 were three-digit numbers rear-projected on a screen before S. In Experiment 3, Ss responded after the illumination of a signal lamp. The Ss indicated their categorizations by pressing one of two buttons on a control panel. Feedback was presented by jewel lamps on the same panel. After every five trials Ss stated their hypotheses regarding the problem's solution. The error results indicated that probability matching did not occur in the MF experiment, but took place in the simple occurrence predicting situation as a result of Ss adopting sequential categorization rules. Another difference between these two experiments was that Ss in the occurrence predicting groups learned at a significantly faster rate. It may be that both differences were due to the large number of irrelevent dimensions in the MF experiment.

In all three experiments mean response latency was longer after an error than after a correct response. Reasoning that hypothesis change (resampling) consumes a certain amount of time, such results tend to support the assumptions of the mathematical models. On the other hand, it is probable that an aversive effect of being told one is "wrong" accounts for at least some of these latency data.

Hypothesis data from the NC experiment make it clear that Ss do not necessarily change their hypotheses immediately after an error. Neither do they always maintain a hypothesis in the absence of an error. Instead, as suggested by the difference in hypothesis alteration rates of the MF and NC experiments, Ss appear to modify their hypotheses only when the overall level of error is uncomfortably high. Although the overall levels of performance differed, the general trend of behaviors and results was similar in all three of the experimental paradigms.