

70-14,747

COLLE, Herbert Anthony, 1943-
THE INFLUENCE OF MEMORY ON THE GAMBLER'S
FALLACY.

University of Washington, Ph.D., 1969
Psychology, experimental

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THE INFLUENCE OF MEMORY ON THE
GAMBLER'S FALLACY

by

HERBERT ANTHONY COLLE

A thesis submitted in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF WASHINGTON

1969

Approved by _____
(Chairman of Supervisory Committee)

Department _____
(Departmental Faculty sponsoring candidate)

Date _____

UNIVERSITY OF WASHINGTON

Date: October 9, 1969

We have carefully read the dissertation entitled The influence of memory
on the Gambler's Fallacy

submitted by
Herbert A. Colle in partial fulfillment of
the requirements of the degree of Doctor of Philosophy
and recommend its acceptance. In support of this recommendation we present the following
joint statement of evaluation to be filed with the dissertation.

Mr. Colle has run three experiments to test a simple memory change hypothesis which was proposed to account for several phenomena that have puzzled students of learning. He suggests that the disappearance of the gambler's fallacy that occurs in the later trials of probability learning experiments and its failure to occur in difficult psychophysical situations are due to a change in the subject's memory for the recent events in the experiment. His experiment demonstrates that after 500 trials on a probability learning experiment subjects are much less likely to learn a sequence constraint that requires a memory of the past 3, 4, 5, or 6 events than they are at the beginning of the experiment. In his second study, he demonstrates a similar failure to remember recent events in a psychophysical recognition experiment. In the third study, subjects were required to predict as in a probability learning study while simultaneously attempting to perform accurately in a task taxing their short term memories. The results of this experiment suggest a mechanism which may account for the memory change.

DISSERTATION READING COMMITTEE:

Richard M. Rose
Danda Y. Teller
Murciuff H. Smith, Jr.

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of Washington I agree that the Library shall make it freely available for inspection. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by my major professor, or, in his absence, by the Director of Libraries. It is understood that any copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Signature _____

Date _____

University of Washington

Abstract

THE INFLUENCE OF MEMORY ON THE

GAMBLER'S FALLACY

by Herbert Anthony Colle

Chairman of Supervisory Committee: Professor Richard M. Rose
Department of Psychology

Recency curves have been used to test several learning models for binary choice experiments in which the sequence of events is generated by a Bernoulli process. The recency curves obtained from the early trials of these experiments have had predominantly negative slopes. However, after about trial 500 the recency curves are either flat or have a slightly positive slope. A recency curve with a slightly positive slope has also been found in the early trials of an analogous two-alternative recognition experiment.

The memory change hypothesis was proposed as an explanation for these differences in the recency curves. Basically, it states that the change from negative to positive recency effects is due to a reduction in the amount of temporary memory the subjects effectively use to analyze the past sequence of events. It was argued that if subjects are less effective in analyzing the immediate past then they should perform less well on a test that requires such analysis.

Two experiments were conducted which were parametric investigations of the analyzing capabilities of subjects in a binary choice experiment (both on the early trials and after 500 trials on a sequence generated by a Bernoulli process) and in a two-alternative recognition experiment. Subjects' analyzing capabilities were tested by introducing rule trials into the

sequence of events. These rule trials, on which 100% correct performance was possible, depended for their solution on a knowledge of the events which occurred on the last k trials, where k took the values 3, 4, 5, or 6. The results supported the memory change hypothesis. Subjects performed less well on the rule trials in the two situations in which positive recency effects are found than they do on the early trials of binary choice experiments where negative recency effects are found.

A third experiment was an exploratory attempt to change the slope of the recency curve on the early trials of a binary choice experiment from a negative to a positive slope by manipulating the amount of short term memory subjects had available for use on the binary choice task. If subjects' temporary analyzing capabilities are strictly a short term memory effect, then the memory change hypothesis predicts that it should be possible to produce positive recency effects by increasing the demands on the short term memory system. The subjects had to make their responses on the binary choice task while holding in memory a list of letters for a short term memory task. The length of the list was varied. None of the recency curves which were obtained showed evidence of being positively sloped. Implications of the experiments and a more specific description of a possible mechanism underlying the subjects' analyzing capabilities were discussed.

ACKNOWLEDGEMENT

There are a number of people whom I would like to thank for their help. Richard M. Rose, my chairman, first interested me in this area of research. He has been a constant source of inspiration and criticism. Suggestions by other members of the committee, especially Moncrieff H. Smith, Jr., and Davida Y. Teller, have made improvements in the design, data analyses, and interpretation of the results, not to mention stylistic clarifications. Richard Bauer read most of an earlier draft of this paper and his comments helped improve by english.

This research was undertaken while I held a National Defense Education Act Title IV Fellowship. In addition, the Department of Psychology provided additional funds for the completion of this research.

The equipment used in the present experiments was obtained from a number of different sources. Moncrieff H. Smith, Jr. generously supplied me with much of the switching, recording, and audio equipment. The audio filter used in experiment II was borrowed from James P. Egan. The Department of Psychology provided the three sound attenuating booths, other audio equipment, and the space to complete this research. Neil Viemeister provided assistance with the design of the audio circuit and was kind enough to measure the outputs of the earphones which were used in experiment II.

I would like to express my sincere appreciation to my wife, Claudine. Not only did she bouy my spirits and see to it that life ran smoothly, but she was also directly responsible for the completion of several hundred hours of data analysis.

TABLE OF CONTENTS

	Page
Introduction	1
The learning of rule sequences	5
Random sequences and the memory change hypothesis	10
Tests of the memory change hypothesis	12
Experiment I	15
Method	15
Subjects	15
Sequences	15
Apparatus	20
Procedure	21
Results and Discussion	22
Experiment II	38
Method	38
Subjects	38
Apparatus	38
Procedure	41
Results and Discussion	42
Experiment III	46
Method	46
Subjects	46
Apparatus	46
Procedure	48
Results and Discussion	53
Final Discussion	61
Bibliography	72
Footnotes	76

LIST OF FIGURES

Figure	Description	Page
1	The proportion of correct responses on run rule trials for the two 3 tuples groups. Each point is based on 15 occurrences of the pair of run rules for each of the 20 subjects in a group.....	23
2	The proportion of correct responses on run rule trials for the six 4, 5, and 6 tuple groups. Each point is based on 15 occurrences of the pair of run rules for each of the 20 subjects in a group.....	24
3	The proportion of correct responses on all-run rule trials for each subject. Each column presents the data from one of the eight experimental groups.....	29
4	The proportion of correct responses on k-span rule trials for the six 4, 5, and 6 tuple groups. Each point is based on 15 occurrences of the pair of k-span rules for each of the 20 subjects in a group.....	31
5	The proportion of correct responses on all k-span rule trials for each subject. Each column presents the data from one of the six experimental groups that had k-span rules in its sequences.....	32
6	The recency curves for the first two blocks of 251 trials for the five groups which received random sequences on these trials. The solid lines show the combined data from all five groups.....	35

LIST OF FIGURES (CONTINUED)

Figure	Description	Page
7	The recency curves for the Random-Random control group in six blocks of 251 trials.....	37
8	A block diagram of the audio circuit used in experiment II.....	39
9	A schematic drawing of the display panel used in experiment II.....	40
10	The proportion of correct responses on run rule trials and on k-span rule trials for the three experimental groups. The data points were computed in the same way as those in experiment I were (see Figs. 1,2, and 4).....	43
11	The proportion of correct responses on all run rule trials and, separately, on all k-span rule trials for each subject..	45
12	The recency curves for the two blocks of 251 trials for the control group;in experiment I. Analogous recency curves from the combined data in experiment I are replotted for comparison.....	47
13	A schematic drawing of the display panel used in experiment III.....	49
14	The order of the trial events in experiment III.....	50
15	The recency curves for group 4 in experiment III, presented separately for the right and left events.....	54

LIST OF FIGURES (CONTINUED)

Figure	Description	Page
16	The recency curves for group 6 in experiment III, presented separately for the right and left events.....	55
17	The recency curves for group 8 in experiment III, presented separately for the right and left events.....	56
18	The recency curves for the control group in experiment III, presented separately for the right and left events.....	57

LIST OF TABLES

Table	Description	Page
1	The k-span rules used for each tuple length and the sampling restrictions they necessitated.....	18
2	The estimated straight line slopes for the recency curves obtained from the four groups in experiment III. The slopes were estimated separately for the left and right events and were computed twice, once using all six data points and again using only the first four data points.....	59

INTRODUCTION

It is a part of our folklore that gamblers display a type of behavior which has come to be known as the gambler's fallacy or sometimes as the Monte Carlo fallacy (Epstein, 1967 pp. 412-413). The behavior arises when people try to predict which of two events will occur next in a sequence generated by a Bernoulli process, that is, a sequence in which each event has a constant probability of occurring on each trial independent of the events on the previous trials. Let y and b denote the two events that are to be predicted and let Y be a prediction that event y will occur and let B be a prediction that event b will occur. The gambler's fallacy refers to a decrease in the probability of making prediction Y (prediction B) as the run length of the y event (b event) increases.

Psychologists have found this type of behavior in binary choice experiments. In the typical binary choice experiment when a warning signal is given the subjects are to predict which of two events will occur. For example, the two events might be the left and right lights on a display panel. Usually the experiment is described to the subjects as a learning or guessing experiment and the subjects are given no prior information about the sequence of events. The gambler's fallacy is also called the negative recency effect by psychologists.

The gambler's fallacy is an anathema for several classes of models which describe how human subjects make predictions. It violates predictions of both decision theoretic models (Edwards, 1955), linear learning models (Bush and Mosteller, 1955), and several stimulus sampling theories (Atkinson and Estes, 1963) which claim that people's predictions are really the result of a very elementary kind of learning process.

Although the decision theoretic models have been applied to it, the binary choice task is not a good situation in which to test them. The decision theoretic models assume that subjects know all the conditions of the experiment (payoffs, probabilities, sampling procedures, etc.) and that the subjects' hypotheses are well-defined. These conditions are not usually met in binary choice experiments. In addition Coombs and Huang (1968) have argued that the decision theoretic models are not really applicable to situations that require repeated decisions. For these reasons decision theoretic models will not be considered further in the present paper.

The path independent linear operator and stimulus sampling theory models predict the exact opposite of the gambler's fallacy. They predict that the probability of making prediction Y (prediction B) will either increase or remain constant as the run length of event y (event b) is increased (Brown and Overall, 1959; Bush and Morlock, 1959). An increase in the probability of making prediction Y (prediction B) as the run length of event y (event b) is increased is called the positive recency effect.

Several explanations have been proposed to account for the failures to find positive recency. Estes (1962, 1964) has argued that the gambler's fallacy (negative recency) is found only (a) when the sequences used were not really generated by a Bernoulli process because of sampling restrictions, or (b) when the data from the early trials of the experiments are analyzed. In fact, these two rules summarize the data from a large number of studies. Studies with sampling restrictions find positive or negative recency depending on whether the sampling technique was biased towards long or short run lengths, respectively (Anderson, 1960; Chapman, 1961; Derks, 1963; Engler, 1958; Feldman, 1959; Hake and Hyman, 1953; Jarvik, 1951; Jones and

Myers, 1966; Nicks, 1959; Rubinstein, 1961). Recency curves from the early trials consistently show negative recency effects when the sequence is generated by a Bernoulli process. Exactly how many trials it takes for the negative recency effect to change to a positive recency effect is not clear, but a safe generalization is that the recency curves are never negative after 500 trials (Brown and Overall, 1959; Bush and Morlock, 1959; Derks, 1962; Edwards, 1961; Estes, 1957; Friedman, Burke, Cole, Keller, Millward and Estes, 1964; Lindman and Edwards, 1961; Nicks, 1959).

Estes (1962, 1964) has argued that the reason that recency curves are negative during the early trials and positive later is that subjects come into experiments with response biases which are extinguished over trials. Although the nature of these hypothesized response biases has not been specified, one property they might be expected to have is to be relatively independent of the task setting. For example, in two-alternative recognition experiments the subjects' task is to make one response if stimulus S_1 is presented and to make the other response if stimulus S_2 is presented. The subject is told via feedback lights which response was the correct one. If subjects in a recognition task setting are given a stimulus S_0 on all trials, then the two-alternative recognition task becomes formally identical to the binary choice task, the same stimulus ensemble is presented on each trial and the only information the subjects can use is the event lights which are ostensibly provided to identify the stimulus presentation. Therefore if simple response biases are responsible for the negative recency effect then the negative recency effect should also occur in recognition experiments. Such a study has been conducted by Rose.¹

In his study subjects were given exactly the same sequence of events that were used by Bush and Morlock (1959) in a binary choice experiment in

which a negative recency effect was found. Rose found that if subjects thought that they were in a recognition experiment then the recency curve for the same sequence was slightly positively sloped; the gambler's fallacy was not in evidence. Why should the subjects' response biases be different in these two studies when the subjects experienced the same sequence of events; why should their guesses be different?

Perhaps the question is better put in the form, "Why are the data from the later trials of prediction experiments similar to the data from the early trials of recognition experiments?" Restle (1961) has suggested that after many trials subjects are bored and that this is responsible for the change over trials. This explanation would not be appropriate for the recognition data unless there is some variable correlated with boredom which is also present in the recognition situation. One such variable might be the amount of effort used by the subjects to analyze the sequence or, alternately, the amount of temporary memory the subjects devote to analyzing the sequence.

In the context of his ideal schema theory Restle (1961) has argued that the gambler's fallacy is due to the way subjects try to analyze a binary sequence. Basically, this point of view stresses that subjects first try to organize sequences in some fashion and the learning process takes place after this organization. If subjects organize recent event presentations and use them as a basis for responding, it is necessary for them to hold some subsequences of recent events temporarily in a memory storage of some kind. Perhaps, after many trials on an unsolvable sequence, subjects either reduce or stop this short term organization or change their method of organizing the sequence.

To facilitate communication the organization of the events on previous

trials will be called a memory code and the process will be called the memory coding process. The brief memory of the memory code will be called the effective temporary memory. The hypothesis that subjects change their memory code or the amount of effective temporary memory they use will be called the memory change hypothesis or explanation.

How could the memory change explanation account for the recognition experiment data? In binary choice experiments the subjects are told, either explicitly or implicitly, that relevant information is contained in the sequence of outcome events. On the other hand, in recognition experiments subjects are led to believe that the only information they are to be using is the stimulus information presented on that trial together with the outcome event on the same trial, the outcome event simply naming the stimulus which was presented. The subjects, if they are obeying instructions, have no reason to organize and remember the past sequence of outcome events; all the information is supposed to be in the stimulus differences. If there are no stimulus differences, then the subjects must guess, just as subjects must guess in binary choice experiments. If the subjects in recognition experiments do not organize the outcome events (or if they impose a reduced amount of organization) then obviously their guesses cannot depend on these organizational processes. Hence their data should not be similar to the binary choice experiment data in which these organizational processes are operating (the early trials) but, if anything, should look like the binary choice experiment data obtained under conditions in which these organizational processes are not operating (the later trials).

The Learning of Rule Sequences

Before discussing possible tests of these ideas, it is necessary to

consider how subjects learn rules inserted into binary sequences. A rule specifies a point in the sequence, following a specified pattern of events, at which one of the two events always occurs. Basically, what a rule does is to state that if the events on the preceding trials meet the specified conditions (whatever they are), then only one of the two events will always occur on the next trial (which of the two events it will be will be specified in the rule). Thus it is possible for a subject to be correct all of the time on these trials. Rules need not apply on every trial. On trials when no rule applies the event which is presented is determined by flipping a coin or by some other equivalent random process.

In the experiments which follow the rules which were used can best be described as tuple rules. That is the rules can most easily be described in terms of the tuple of events which occurred on the preceding k trials. A k -tuple of events specifies the events which occur on the k trials from trial n through trial $n + k - 1$ and the order in which the events occurred. Since this specification is in terms of the preceding k trials the analysis is done in an overlapping manner, looking at k adjacent trials while moving ahead one trial at a time. For example, for a tuple of length four, if the pattern of events preceding trial n was $ybyy$ then only two possible 4-tuples can precede trial $n + 1$, $bbyb$ or $bbyy$. An example of a tuple rule is: if whenever the pattern of events on trials n to $n+3$ is $bbyb$ and on trial $n + 4$ the event y always occurs, then trial $n + 4$ is the rule trial. The rule specifies that event pattern $bbyb$ is always followed by event y .

Following suggestions made by Hake and Hyman (1953), Goodnow (1955) and Nicks (1959), several models have claimed that subjects do not use all possible combinations of outcome events as cues, but instead they use only the runs of homogeneous events as cues (Gambino and Myers, 1966; Restle,

1961; Greeno cited by Restle, 1966). This memory code will be called the run length code. The run length code is a variable length code, that is, the number of previous events which are used as cues is not a constant, instead it depends on the length of the ongoing run of events. For example if, starting on trial n , the pattern of events which is presented is byyyby then the pattern of events which is used as a cue on trial $n + 2$ is by, the event pattern which is used as a cue on trial $n + 3$ is byy, the event pattern which is used as a cue on trial $n + 4$ is byyy, the event pattern which is used as a cue on trial $n + 5$ is yb, and so on. Thus a subject using the run length code does not analyze all possible patterns of events and, in addition, he uses a variable length coding process.

According to an opposing point of view, subjects do not impose this organization, but instead they hold in memory all the events which occurred on the last k trials (Burke and Estes, 1957; Restle, 1961). This second possible memory code (or lack of one) will be called the k -span code. The k -span code is both non-selective (all possible tuples of events are used as cues), and also is fixed in length (subjects are described as analyzing a constant number of the immediately preceding events). For example, if the pattern of events which is presented starting on trial n , is yyybybb and if the subjects are analyzing only the preceding three trials, then the pattern of events used as a cue on trial $n + 3$ is yyy, the event pattern used as a cue on trial $n + 4$ is yyy, the event pattern used as a cue on trial $n + 5$ is yyb, and the event pattern used as a cue on trial $n + 6$ is yby.

The difference between these two types of models is an important one for the memory change hypothesis, only because if memory changes it is necessary to know how to measure the change and also because if it is

possible to demonstrate that subjects use an organizational system when there are rules in the sequence then it is plausible that they use the same system when there are no rules in the sequence, especially since subjects are not very often told that the sequence contains no rules.

These two claims were put to a direct test by Rose and Vitz (1966). They presented subjects with sequences that had two different kinds of rules in them, run rules and k-span rules. Both kinds of rules depended on the subjects remembering at least the last k trials. Not every trial was a rule trial; the events which were presented on non-rule trials were chosen randomly. There are only two possible run rules in each sequence of events. For the 4 tuple rules which were used, the two run rules are: the event pattern ybbb (event pattern byyy) is always followed by the event b (event y) and the event pattern bbbb (event pattern yyyy) is always followed by the event y (event b). An alternative description of the two run rules is that runs of lengths three never occurred and runs of length four were the longest run lengths that ever occurred. All the other possible rules were grouped into a separate category called the k-span rules. Out of the six possible k-span rules² the two that were used are: event pattern ybyb (event pattern byby) is always followed by event b (event y) and event pattern bbyb (event pattern yyby) is always followed by event y (event b).

If subjects organize by run lengths then the k-span rules can be learned only after they are reinterpreted into more complicated groups of run rules, and therefore the k-span rules should be harder to learn. On the other hand, if subjects impose no organization on the sequence (the k-span code) then all the rules should be learned equally fast. Rose and Vitz found that the run rules were learned much faster than the k-span rules.

In fact, there was little evidence that the subjects learned the k-span rules. Thus it is clear that subjects are more sensitive to runs of events than they are to other patterns.

Another related line of research is the research on fixed sequences of binary events that are repeated endlessly until the subjects reach a criterion of learning. In these sequences every trial is a rule trial. What kind of organization do subjects impose on these lists? Keller (1963) found that the difficulty of learning lists was unrelated to the number of events in the list (list length) but was directly proportional to the number of runs in the list, again singling out runs as salient cues. In addition, both Restle (1967) and Vitz and Todd (1967) have independently presented identical models for the learning of fixed lists, although Restle extended his version further. These models again assume that subjects use the run length coding process on sequences and they specify the positions in the lists at which subjects should and do make more errors because of this organization.

To summarize, the above evidence suggests that subjects, given sequences with rules in them, organize the sequences into run lengths. This run length code is remembered temporarily so that it can be changed appropriately on the next trial. A more permanent learning process, then, acts on the coded information.

There are several possible ways that subjects could accomplish the run length code. One way is to count the number of consecutive repetitions of each event, temporarily remembering the count so that it can be increased or started anew on the next trial. With this system the subjects then could learn the counts in a way similar to that used by subjects when they learn paired associate or serial lists. Some support comes from a study by

Warden (1924). He had subjects learn stylus mazes. Afterwards he asked them what techniques they had used to learn the mazes. Out of 60 subjects, 25 reported using verbal techniques, 18 reported using visual imagery to determine the spatial characteristics of the maze, and 17 reported just using motor learning. Even though they had spatial cues available, an estimated 40% of the subjects used verbal techniques which consisted mainly of counting. The binary choice experiments reported above minimize the possibility of using spatial cues; they are more like the temporal mazes used with animals. Thus the number of subjects using verbal techniques might be expected to be even higher than the 40% found by Warden.

Another way subjects might accomplish the run lengths code is to use a rhythm to encode the sequence, storing the subrhythms temporarily each trial. The rhythm could compress in time the sequence to be remembered or it could break it up into smaller, more manageable chunks. However this rhythm method (no pun intended) seems unlikely because of the relatively long time between the events and also, because the flexibility of rhythms favors a k-span coding process. In any case at the present time it will not be necessary to analyze the encoding mechanism in detail. It is only necessary to assume that subjects analyze sequences into run lengths and that they must store these runs in a temporary memory system.

Random Sequences and the Memory Change Hypothesis

The memory change hypothesis can now be specified more exactly. The occurrence of the gambler's fallacy in binary choice experiments can be attributed to the subjects' organization of the sequence into run lengths. Positive recency is found on the later trials of prediction experiments because the run length organization is discontinued or rendered less

effective. Either the subjects use a different organization (e.g., k-span) or they effectively reduce their memories thereby impairing the run length organization. Subjects in the recognition experiment setting either do not use the run length organization, or their memories for the past events are reduced to the extent that the run length organization is impaired. It seems unlikely that subjects in recognition experiments would use the more difficult k-span organization so that the positive recency effect, there, probably would be due to a memory reduction. Because of the similar results obtained in the two situations, it might also be expected that the positive recency on the later trials of the prediction experiments also is due to a memory reduction and not to a shift to the k-span organization. Notice that if subjects completely eliminate their short term analysis, then the memory change hypothesis is not incompatible with Estes' (1962, 1964) suggestions that the data from the later trials of prediction experiments can be described adequately by the path-independent models.

To illustrate what is meant by saying that subjects discontinue using the run length code or by saying that subjects effectively reduce their temporary memory usage, assume for convenience that subjects code the sequence by counting successive events. One way subjects could discontinue the organization is just to stop counting. One way in which a subject could effectively reduce his temporary memory without stopping counting is to increase his probability of forgetting what the count was on the previous trial or to increase his probability of forgetting to change the count. In either case as the number of trials back on which the rule depends is increased, and the higher the probability of forgetting the more rapid the increase. Restle's (1961) suggestion that subjects on the later trials of binary choice experiments are "bored" is not incompatible with the

memory change hypothesis if boredom implies that subjects are less attentive or have a greater probability of forgetting, etc.

Tests of the Memory Change Hypothesis

The present paper reports three experiments which were conducted to investigate the plausibility of the memory change hypothesis. Two of the experiments were parametric investigations of the amount of effective temporary memory being used by subjects in the two positive recency situations, the later trials of binary choice experiments and recognition experiments. The memory change hypothesis predicts that subjects will perform less well on the run rules in these two situations. In addition a check was made to see if subjects in these situations switched to using the k-span code instead of the run length code. The third experiment was an exploratory attempt to change the slope of the recency curve from negative to positive by varying the amount of short term memory subjects had available for use in a binary choice experiment. If the subjects' effective temporary memory for past events is strictly a short term memory effect and if the change from negative to positive recency effects is due to a reduction in the subjects' effective temporary memory, then the memory change hypothesis would predict that positive recency effects would be found when the amount of short term memory used by subjects in binary choice experiments is reduced. However, if the subjects' effective temporary memory is not strictly a short term memory phenomena then the prediction the memory change hypothesis makes would depend on the exact mechanism underlying this effective memory.

In experiment I four groups of subjects were given rule sequences similar to those used by Rose and Vitz (1966). Another four groups

received the same rule sequences after first receiving 500 trials on a sequence generated by a Bernoulli process. A third control group received the same number of total trials, all of them being generated by a Bernoulli process. Each of the experimental groups received a rule sequence that required the subjects to have an effective memory for at least the last k trials, where k took one of the values 3, 4, 5, or 6 for each experimental group. The rules were selected so that both events had an equal probability of following any pattern of events of length $k-1$ or less, that is the probability was $\frac{1}{2}$ that either event would occur following any tuple of events based on less than the last k trials. Thus to the extent that the subjects' effective memory is reduced on later trials, they should have more difficulty on the longer rules compared to subjects who get the rule trials initially. Each sequence contained instances of the two run rules and instances of two different k -span rules.

Experiment II was a similar parametric analysis of a two-alternative recognition experiment. Subjects were instructed to report whether a 1000 Hz sine wave was the louder or quieter one and were ostensibly given feedback to help them identify the tones. For the first 50 trials the subjects received two tones with an intensity difference set so that the subjects were correct about 70% of the time. Only one of the two tones was presented on each trial, the two tones being presented in an irregular order. On trial 51 and following the subjects were always presented a tone intensity half way between the previous two intensities, and the rule sequences used in experiment I were presented. Four groups were used. Three groups received one of the rule sequences that depended on either the last 3, 4, or 5 trials. The fourth group was a control group which received the random sequence used in the first experiment. The fourth group is a replication

of Rose's recognition experiment¹ with a number of modifications.

In experiment III an exploratory attempt was made to change the slope of the recency curve from negative to positive by manipulating the subjects' short term memory. The subjects were instructed that their task was to try to do two tasks simultaneously. One task was a variation of the short term probe memory task used originally by Waugh and Norman (1965); the other task was a binary choice task. In the probe memory task subjects received a list of letters to remember. Later one of the letters from the list was presented (the probe) and the subjects' task was to report the letter which followed the probe letter in the original list. The two tasks were combined in a way that forced the subjects to make their binary choice task response while they held a list of letters in short term memory. The two tasks were combined in the following way: (a) A list of letters for the probe memory task was presented. (b) Following the list a warning light was presented and it signaled the subjects to make their binary choice task prediction. (c) The binary choice task outcome event followed immediately after the subjects' predictions. (d) Next another warning light was presented, the probe letter was presented, and the subjects were given an opportunity to report the correct answer for the probe memory task. After the report interval a different list of letters was presented and the sequence of trial events continued in the same way.

The subjects were instructed to attend mainly to the probe letter task, and to merely guess at the binary choice task unless they could try to be correct more often without impairing their probe memory task performance. Notice that for the subjects to be correct on the probe letter task, they had to hold the original list in memory while they responded to the binary choice task. Four groups of subjects were used. Three groups received

lists of either 4, 6, or 8 letters; the presentation time was held constant by varying the presentation rate. A fourth control group received no list at all. If the memory change hypothesis is correct and if the subjects' effective memory for past events is strictly a short term memory effect, then as the amount of effective memory is decreased by increasing the difficulty of the secondary probe memory task, the gambler's fallacy should be replaced by a positive recency effect.

EXPERIMENT I: METHOD

Subjects

The subjects were 180 undergraduates at the University of Washington. The subjects participated to fulfill part of a course requirement.

Sequences

Two kinds of sequences were used, random sequences and sequences containing rule trials. In all of the sequences the y event and the b event occurred equally often. Nicks' (1959) procedure was used to generate the random sequences. The expected number of runs of each length for samples of 500 trials was calculated, and then a population of run lengths was constructed by making the frequency of each run length approximately equal to the expected frequency. Since both events occurred equally often the distributions of run lengths for each event are identical. The distribution was truncated so that no run lengths greater than eight occurred. The probability of obtaining a run length of either event greater than or equal to nine is less than 0.004. The two distributions of run lengths (one for each event) were then divided approximately in half so that blocks of 250

trials would have approximately the same run structure. Finally, the run lengths were sampled without replacement from the populations. This was done by drawing a sample from the population for one of the events, drawing the next sample from the population for the other event; drawing a sample from the population for the first event again, and so on until both populations were exhausted. This procedure was used twice to generate two different random sequences of 502 trials, which are referred to below as random sequences one and two. This process generates sequences in which $\frac{1}{2}$ the runs are of length 1, $\frac{1}{4}$ are of length 2, $\frac{1}{8}$ are of length 3, $\frac{1}{16}$ are of length 4, and so on until the truncation point is reached.

The sequences containing rule trials were generated by the same procedure except that additional constraints were placed on the sequence of run lengths. The run rules and the k-span rules each required different constraints. The two run rules in each sequence were generated by the following device. For the n-tuple run rules, runs of length $n-1$, n , and $m > n$ were all made into runs of length n . This was accomplished simply by sampling from the populations which were constructed for the random sequences, but every time a run of length $n-1$ or greater was sampled a run of length n was entered in the sequence. These restrictions generate sequences with two different run rules. Run rule one states that runs of length $n-1$ never occur, and run rule two states that there are no runs longer than n .

For the 3-tuple rules, run rule one states that runs of length 2 never occur and run rule two states that there are no runs longer than 3. In terms of tuples of events run rule one states that event pattern ybb will be followed by event b and run rule two states that event pattern bbb will be followed by event y. Here, and in the following presentation, rules obtained merely by interchanging the y and b events will be considered equivalent.

For example, the rule: event pattern ybb is followed by event b, will be considered equivalent to the rule: event pattern byy is followed by event y. For the 3 tuple rules the restrictions generate sequences in which $\frac{1}{2}$ of the runs are of length 1, and $\frac{1}{2}$ of the runs are of length 3.

For the 4-tuple rules run rule one states that runs of length 3 never occur and run rule two states that there are no runs longer than 4. In terms of tuples of events run rule one states that event pattern ybbb will be followed by event b and run rule two states that event pattern bbbb will be followed by event y. These restrictions generate sequences in which $\frac{1}{2}$ of the runs are of length 1, $\frac{1}{4}$ are of length 2, and $\frac{1}{4}$ are of length 4.

For the 5-tuple rules, run rule one states that runs of length 4 never occur and run rule two states that there are no runs longer than 5. In terms of tuples of events run rule one states that event pattern ybbbb will be followed by event b and run rule two states that event pattern bbbbb will be followed by event y. These restrictions generate sequences in which $\frac{1}{2}$ of the runs are of length 1, $\frac{1}{4}$ are of length 2, $1/8$ are of length 3, and $1/8$ are of length 5.

For the 6-tuple rules, run rule one states that runs of length 5 never occur and run rule two states that there are no runs longer than 6. In terms of tuples of events run rule one states that event pattern ybbbbb will be followed by event b and run rule two states that event pattern bbbbbb will be followed by event y. These restrictions generate sequences in which $\frac{1}{2}$ of the runs are of length 1, $\frac{1}{4}$ are of length 2, $1/8$ are of length 3, $1/16$ are of length 4, and $1/16$ are of length 6.

Table 1 shows for each tuple length the k-span rules which were used and also the method used to generate the k-span rules, that is the restrictions placed on the sequence of run lengths. For each n-tuple sequence,

Table 1

The k-span rules used for each tuple length and the sampling restrictions they necessitated.

Rule Length	k-span Rules	Sampling Restrictions on the Run Lengths
3	None	
4	(1) $yyby \rightarrow b$ $bbyb \rightarrow y$	(1) $2-1 \rightarrow 1$ and $4-1 \rightarrow 1$
	(2) $ybyb \rightarrow b$ $byby \rightarrow y$	(2) $2-1-1 \rightarrow 2$ or 4 and $4-1-1 \rightarrow 2$ or 4
5	(1) $yybby \rightarrow b$ $bbbyb \rightarrow y$	(1) $2-2 \rightarrow 1$ and $3-2 \rightarrow 1$ and $5-2 \rightarrow 1$
	(2) $bybby \rightarrow y$ $ybyyb \rightarrow b$	(2) $1-2 \rightarrow 2$ or 3 or 5
6	(1) $ybyyby \rightarrow b$ $bybbyb \rightarrow y$	(1) $1-2-1 \rightarrow 1$
	(2) $bbyyby \rightarrow y$ $yybbyb \rightarrow b$	(2) $2-2-1 \rightarrow 2$ or 3 or 4 or 6 and $3-2-1 \rightarrow 2$ or 3 or 4 or 6 and $4-2-1 \rightarrow 2$ or 3 or 4 or 6 and $6-2-1 \rightarrow 2$ or 3 or 4 or 6

out of a possible $2^{n-1}-2$ k-span rules two were included in the sequence.² All the other tuples, except the run rule tuples, were followed by each of the two events with probability $\frac{1}{2}$. For example, the two 4 tuple k-span rules were: (a) If the pattern of events on the last four trials was yyby (or bbyb) then the b event (y event) will always occur on the next trial. (b) If the pattern of events on the last four trials was ybyb (or byby) then the b event (y event) will always occur on the next trial. These two rules are summarized in column one by bbyb \rightarrow y, yyby \rightarrow b and byby \rightarrow y, ybyb \rightarrow b. The two restrictions which were used to generate these rules are: (a) If a run of length two or four is followed by a run of length one, then a run of length one is always selected next. This was accomplished by drawing run lengths from the population described above. For example, if a run of length 2 was drawn and if on the next draw a run of length 1 was drawn then draws were made and replaced until a run of length 1 was drawn. (b) If a run of length two or four is followed by two runs of length one then a run of length two or four is always selected next. These two sampling restrictions are summarized in column two by 2-1 \rightarrow 1, 4-1 \rightarrow 1 and 2-1-1 \rightarrow 2 or 4, 4-1-1 \rightarrow 2 or 4. The other k-span rules can be interpreted similarly. There were no 3 tuple k-span rules because it is impossible to include them in the same sequence with the run rules without getting a sequence in which some of the event patterns never occur. This sampling procedure was used to generate two separate sequences of 992 trials (3 blocks of 256 trials and 1 block of 224 trials) for each rule length (3, 4, 5, and 6 tuple rule sequences one and two).

The entire set of four rules, the subset consisting of the two run rules, and the subset consisting of the two k-span rules, which were included in each sequence (except the 3 tuple sequences which contained only run

rules), have the property that the unconditional probability of each event and the probabilities of each event conditional on any tuple of events which is shorter than the rule tuple is equal to $\frac{1}{2}$. Thus by evaluating performance on both rules of a pair, effects of bias which are independent of the events or which depend upon tuples which are shorter than the rule tuples can be eliminated. This is not true when the rules are evaluated separately. For example, if run rule one is evaluated by itself then a simple tendency to predict that the event which occurred on the previous trial will be repeated will lead to performance above the chance level. However this same tendency will lead to performance below the chance level on run rule two. Thus by considering performance on both run rules together and on both k-span rules together, the effects of these biases is diminished.

Apparatus

The apparatus consisted of a long table divided into five booths by curtains. Each booth was equipped with a pair of keys which the subjects used to make their predictions. All five booths faced a single 19 in. wide and 14.5 in. high display panel, located about 10 ft. across the room from the subjects. A yellow warning light was located 9.5 in. from each end of the panel. Two red event lights were located slightly below the warning light and 5.5 in. from the center of the panel.

The sequence of trial events was programmed to run automatically. The programming equipment was located in the same room as the subjects but it was activated only after all the subjects had responded. This procedure eliminated possible extraneous cues from the programming equipment.

Procedure

The subjects were tested in groups of five or fewer. They were instructed that when the yellow warning light came on they were to predict which of the two event lights would come on. The left key was to be pressed to predict the occurrence of the left light and the right key was to be pressed to predict the occurrence of the right light. The trials were group paced; the appropriate event light came on 1.4 sec. after all the subjects responded. It stayed on for $\frac{1}{2}$ sec. and was followed by a $\frac{1}{2}$ sec. interval before the warning light came on again.

Besides specifying the details of the procedure, the instructions contained the information that the experiment was a learning experiment and that the subjects were to predict the event which would occur. The subjects were told nothing about the sequence of events. They were told that there was no group interaction and were asked to try hard because the experiment was a test of the limitations of the human learning process.

Nine groups of 20 subjects were tested. Each group received a different sequence of event lights. Four groups (Rule-only groups) were given 992 trials on one of the four rule sequences. Another four groups (Random-rule groups) were given 502 trials on one of the random sequences followed without any break or cues by 992 trials on one of the four rule sequences. A control group (Random-random group) was given 1506 trials on the random sequences. A maximum time limit of 100 min. was established.

Within each of the rule-only groups $\frac{1}{2}$ of the subjects were tested on rule sequence one and $\frac{1}{2}$ were tested on rule sequence two. Within each of the random-rule groups $\frac{1}{2}$ of the subjects were tested on random sequence one and rule sequence one, and $\frac{1}{2}$ were tested on random sequence two and rule

sequence two. Within the random-random group $\frac{1}{2}$ of the subjects received 502 trials on random sequence one, 502 trials on random sequence two, and then 502 trials on random sequence one again. The other $\frac{1}{2}$ of the subjects in the random-random group received 502 trials on random sequence two, 502 trials on random sequence one, and 502 trials on random sequence two again. Thus each of the nine groups had two subgroups which received independently generated sequences. Each of these subgroups were balanced so that for $\frac{1}{2}$ of the subjects in each subgroup the sequence started with the left event and for the other $\frac{1}{2}$ of the subjects it started with the right event, the right and left events being interchanged.

Two subjects in the experimental groups were discarded. One subject was discarded before her data were analysed because she reported taking heavy medication before the experimental session and being unable to concentrate because of its effect. She was replaced by another subject. The other subject was discarded from group 3 because he made correct responses only 69 of the 480 rule trials. The probability of getting only 69 or fewer correct responses by chance alone is much less than 0.001. Thus the subject was discarded for either misunderstanding the instructions or for deliberately choosing the incorrect response. His data were not replaced so that group 3 consists of only 19 subjects.

EXPERIMENT I: RESULTS AND DISCUSSION

Figures 1 and 2 present the run rule learning curves for the eight experimental groups. The data from the two run rules in each sequence were combined to eliminate effects of simple strategies and biases. Strategies such as simple and double alternation, response repetition, or any strategy based on a tuple of events which is shorter than the rule tuple will lead

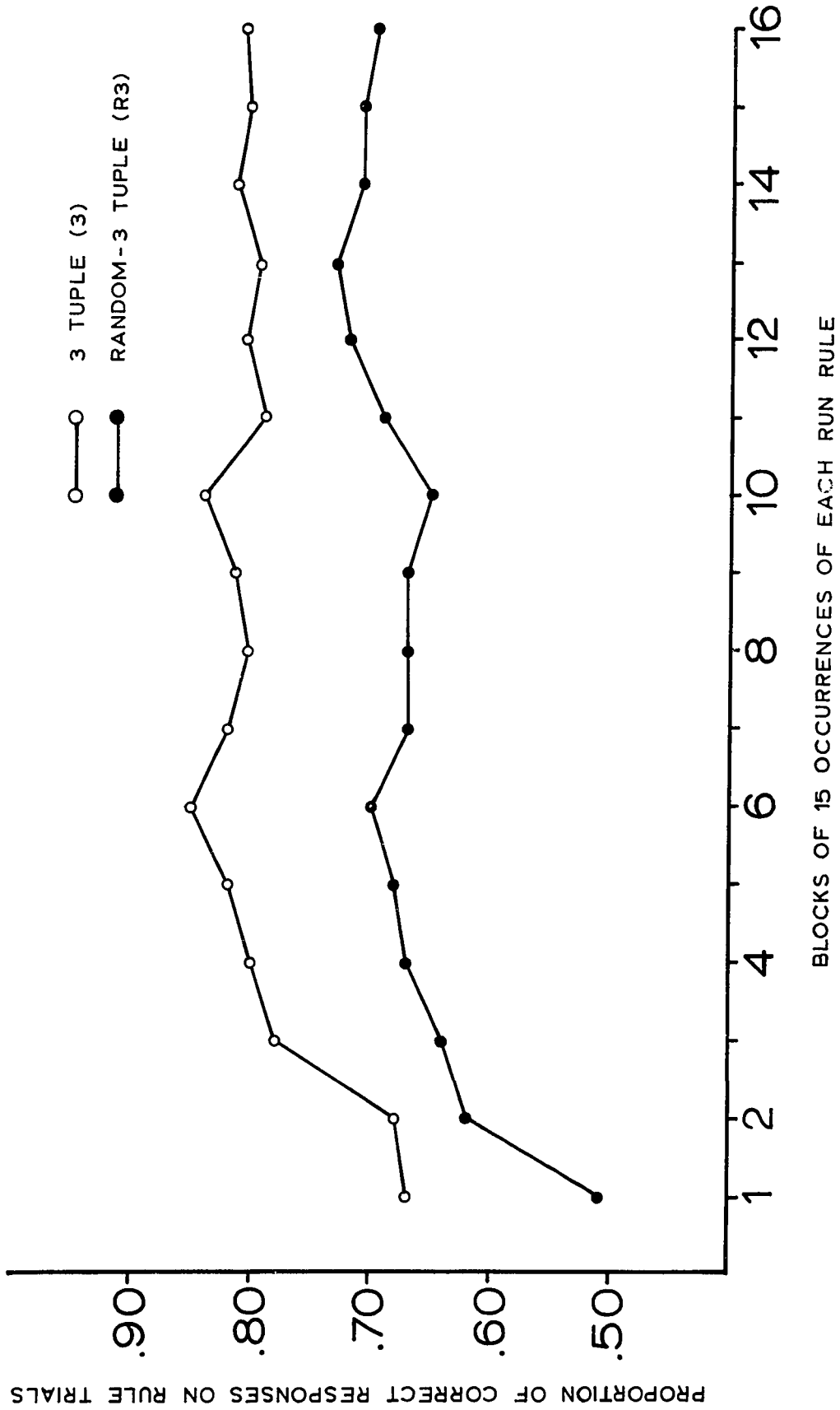


Figure 1. The proportion of correct responses on run rule trials for the two 3 tuple groups. Each point is based on 15 occurrences of the pair of run rules for each of the 20 subjects in a group.

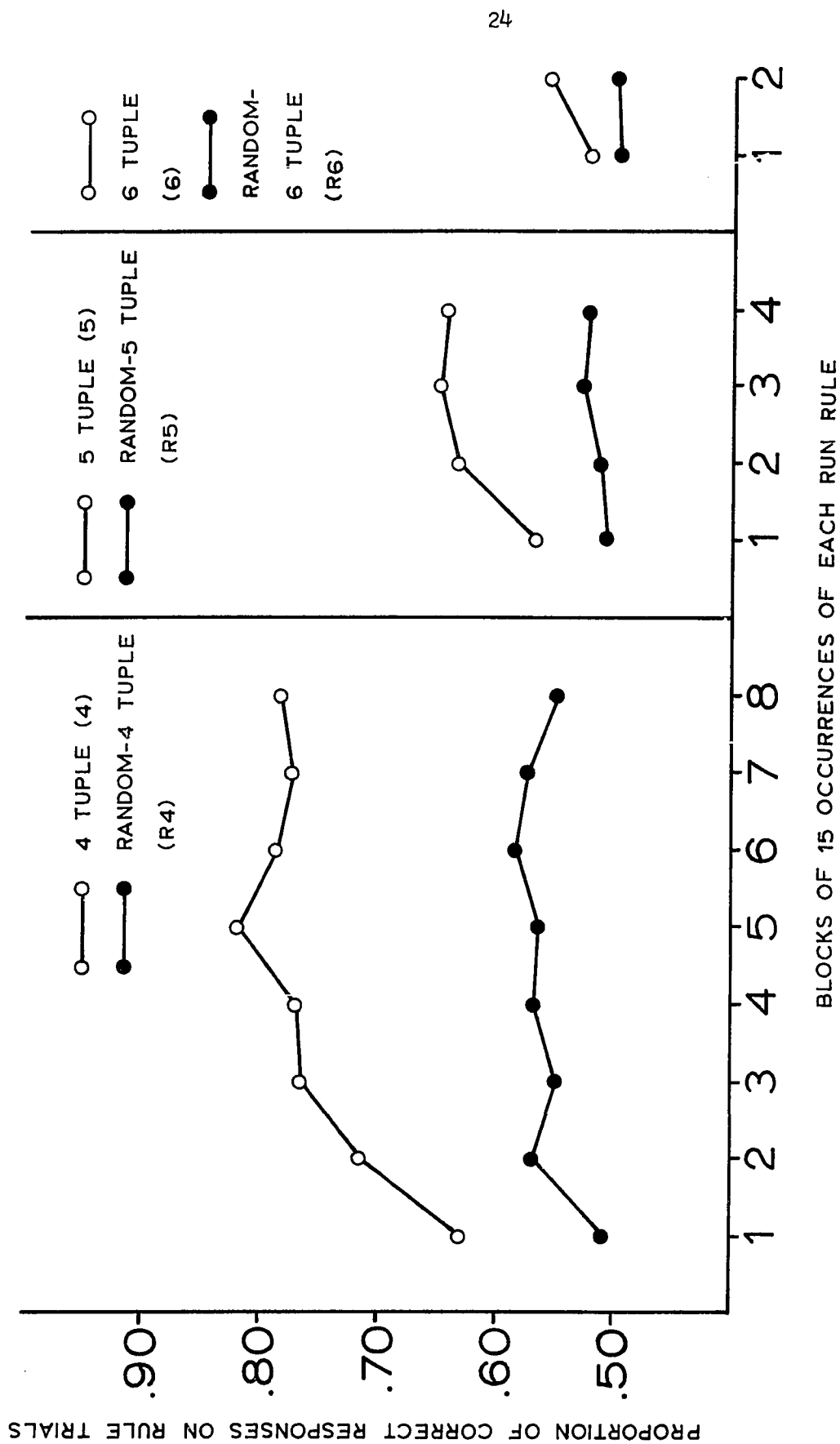


Figure 2. The proportion of correct responses on run rule trials for the six 4, 5, and 6 tuple groups. Each point is based on 15 occurrences of the pair of run rules for each of the 20 subjects in a group.

to chance performance when the data from the two rules are combined (see the end of the sequences section). Each block of trials is based on 15 occurrences of each of the two run rules for each of the 20 subjects in the group. Thus each point is based on 600 responses and it shows the overall proportion of correct responses made by the 20 subjects on the 90 rule trials. To evaluate the curves note that, assuming binomial variability (i.e., one standard deviation equals $(pq/n)^{\frac{1}{2}}$), the maximum size of a 95% confidence interval about each point is ± 0.04 for an N of 600.³ This maximum is obtained when the proportion correct is 0.50; for proportions greater than 0.50 the confidence intervals are somewhat smaller. For example, when the proportion correct is 0.80 the 95% confidence interval is reduced by a factor of 0.8; the interval is about ± 0.03 . These confidence intervals will be used to make statistical comparisons in the following paragraphs.

First of all, notice that the learning curves for the groups receiving 3, 4, and 5 tuple rules appear to have reached asymptote. No group reached the maximum possible correct performance of 100%. Groups 3 and 4 reached their asymptote by the third block of trials, groups R4 and 5 reached theirs by the second block of trials, and group R3 reached its by the fourth block of trials. Group 6 was not given enough rule trials to determine an asymptote, but on the basis of the other groups' performance the group 6 asymptote is probably below 0.60.

All of the groups which received the rules at the beginning of the session performed significantly above the chance level. Groups 3 and 4 have asymptotes at about 0.80. Group 5 has an asymptote at about 0.65. By the second block of trials group 6 reached 0.56. Of the groups that received the rules after the random trials only groups R3 and R4 have

curves that are significantly greater than chance. Group R3 has an asymptote at 0.70 and group R4 has one at 0.57.

The learning curves for the groups which received the random sequences before getting the rules (groups R3, R4, R5, and R6) are all significantly below the corresponding learning curves for the groups which received the rules at the beginning of the session (groups 3, 4, 5 and 6). The difference between groups 3 and R3 is about 0.10, between groups 4 and R4 is about 0.20, and between groups 5 and R5 is about 0.10. The difference between groups 6 and R6 for the second block of trials is 0.06.

It is clear, then, that subjects who were first exposed to 500 random trials performed less well than subjects who received the rules at the beginning of the session. In addition, only the R3 and R4 groups showed evidence of having learned the rules; groups R5 and R6 did not perform significantly above the chance level. A comparison of the 3 tuple and the 4 tuple groups showed that although groups 3 and 4 do not differ, groups R3 and R4 are significantly different. The performance on the 4 tuple rules, which required a greater effective memory, was impaired more by the random trials than the performance on the 3 tuple rules was. It is harder to make such comparative statements about the 5 and 6 tuple groups because the performance of the groups who received the rules at the beginning of the session (groups 5 and 6) is significantly below the performance of the corresponding 3 and 4 tuple groups. In addition, the performance of the random groups is bounded below by 0.50. Therefore quantitative comparisons of the magnitude of the real differences between groups 5 and R5 and between 6 and R6 cannot be made. It is clear, however, that groups R5 and R6 show no evidence of learning.

There are at least two explanations for the decreases in performance

that the groups which received the rules at the beginning of the session showed with increases in the tuple length. The first is simply that the longer tuple lengths are harder than the shorter ones. This assumption is implicit in the concept, "effective memory" used in the present paper, although most of the theories which assume that subjects use the run length code do not make this assumption: they would predict asymptotically perfect performance on all run rules. The second possible explanation for the decrease is that the "trials effect" influenced the performance of the groups with longer tuple lengths. This is possible because it takes twice as many trials to present m occurrences of a tuple rule of length k than it takes to present m occurrences of a tuple rule of length $k-1$. Thus the same process that reduces subjects' effective memory during random trials may take place on the random trials between the rule occurrences in the rule sequences. A similar decrease in performance on longer tuple rules may be found in Restle's (1966) data.

An inspection of the group curves and the individual subjects' protocols revealed that the data from the eight experimental groups can be summarized fairly accurately by two summary statements. The first statement is that regardless of when a run rule is introduced, either individual subjects improve their performance rather rapidly until some individual asymptote is reached or they show no improvement in performance. The second statement is that the greater the number of trials received before the first m occurrences of the run rules the less likely it is that an individual subject will show any improvement in performance. A corollary of the second statement is that some subjects in all the groups may never show any improvement. The second statement is based not only on the

difference between the groups who received the rules before and after 500 random trials, but it is also based on the reduction in subjects' performance on the longer rule tuples. It was argued above that one of the possible reasons for this reduction was the fact that it takes twice as many trials to present m occurrences of a tuple rule of length k than it takes to present m occurrences of a tuple rule of length $k-1$.

Figure 3 shows a summary of the individual subjects' performance. Each data point presents the total proportion of correct responses on all the run rule trials for an individual subject. The arrows point to the median subject's performance. The horizontal bars show the point c at which the binomial probability of getting a proportion greater than or equal to c is 0.05. The above descriptions of the behavior of the group curves can be seen again in the distributions of the individual data points. In addition Figure 3 illustrates to some extent the statements summarized above. The clearest example is the 3 tuple rule distributions. Group 3's distribution is skewed downward; the majority of the subjects performed very well with a few subjects who show no increase in performance. On the other hand, group R3's distribution appears to be bimodal; most of the subjects either performed very well or performed at the chance level. Four subjects' protocols do not show the rapid improvement implied by the first summary statement. Two subjects in group 3 showed a significant improvement in performance first in block 5 and two other subjects in group R3 also showed a significant improvement late in the sequence, one in block 7 and one in block 12. Although group 4's data points are uniformly distributed about the median and show no support for the first summary statement, inspection of the individual subjects' protocols revealed that the data points in the 0.60-0.75 range were generated by subjects who consistently

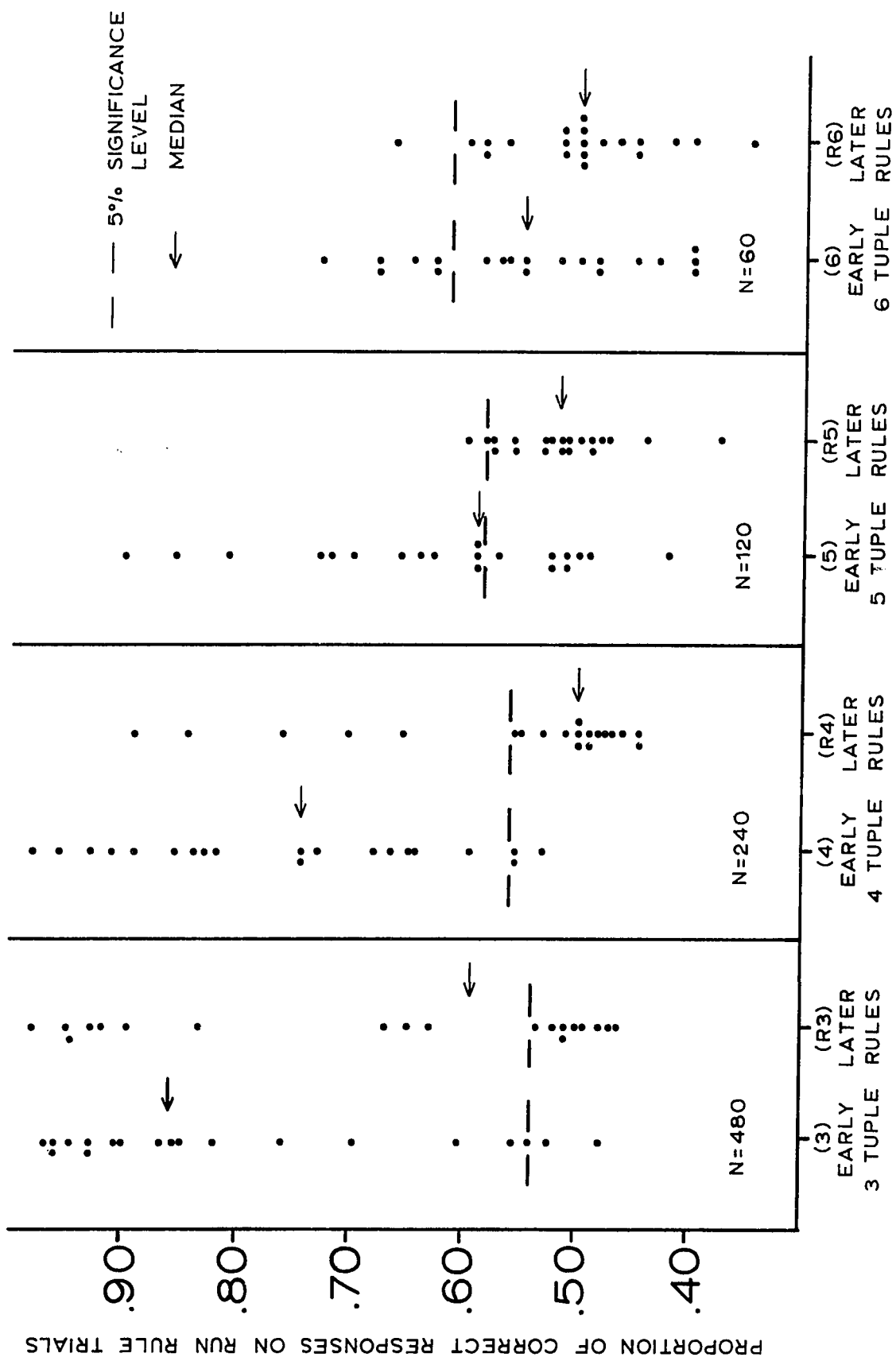


Figure 3. The proportion of correct responses on all run rule trials for each subject. Each column presents the data from one of the eight experimental groups.

get more than 15 out of the 30 rule occurrences per block correct, but perform below the 23-30 range.

Figure 4 presents the learning curves for the k-span rules and Figure 5 presents the individual subjects' data. Each block of trials is based on 15 occurrences of each of the two k-span rules for each of the 20 subjects in a group. The curves in Fig. 4 can be evaluated in the same way as the curves in Figs. 1 and 2 were. All the 5 and 6 tuple groups show no evidence of an increase in performance; all four curves are not significantly different from chance performance. The individual data points in Fig. 5 support this conclusion. Both of the 4 tuple groups show evidence of an increase in performance significantly greater than chance performance. Note, however, that group R4's curve paradoxically drops back to chance performance on the later trials and group 4's curve appears to be still increasing. A comparison of group 4 and R4 shows that they differ only after block 5 when group 4 increases 0.05 and group R4 decreases at least 0.05. Thus unlike its effect on the run rules, the random trials seem to have influenced the subjects' performance on the k-span rules only later in the rule sequence. The strange behavior of these two curves precludes putting much confidence in these results.

A comparison of the run and k-span rule learning curves for group 4 (see Figs. 2 and 4) showed that the run rule learning curve was significantly above the k-span rule learning curve for every block of trials. The differences between the curves was always at least 0.10. In addition, for each subject the overall proportion of correct responses he made on run rule trials was compared with the overall proportion of correct responses he made on k-span rule trials. Of the 20 subjects in group 4, 19 of them had a higher proportion of correct responses on run rule trials than they

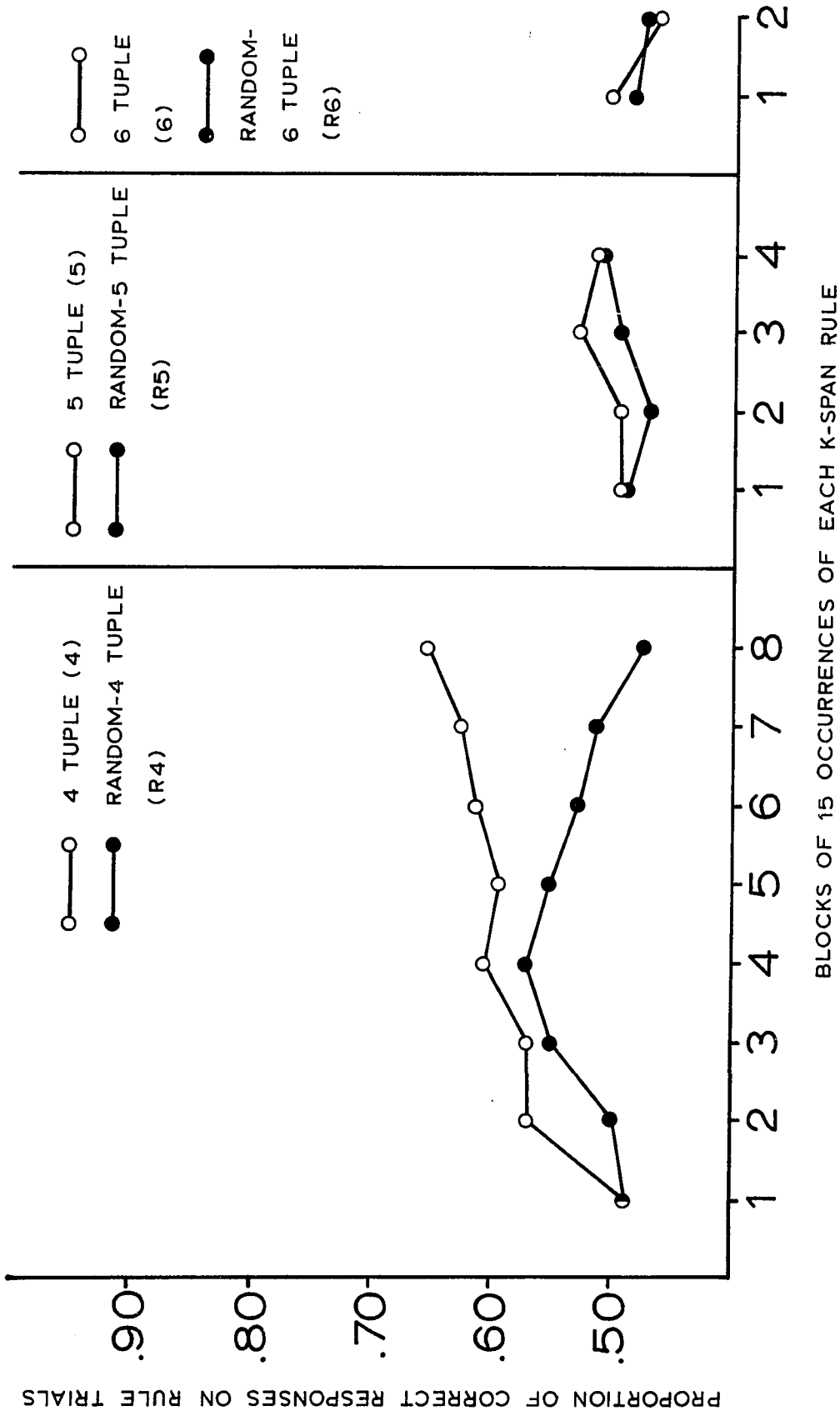


Figure 4. The proportion of correct responses on k-span rule trials for the six 4, 5, and 6 tuple groups. Each point is based on 15 occurrences of the pair of k-span rules for each of the 20 subjects in a group.

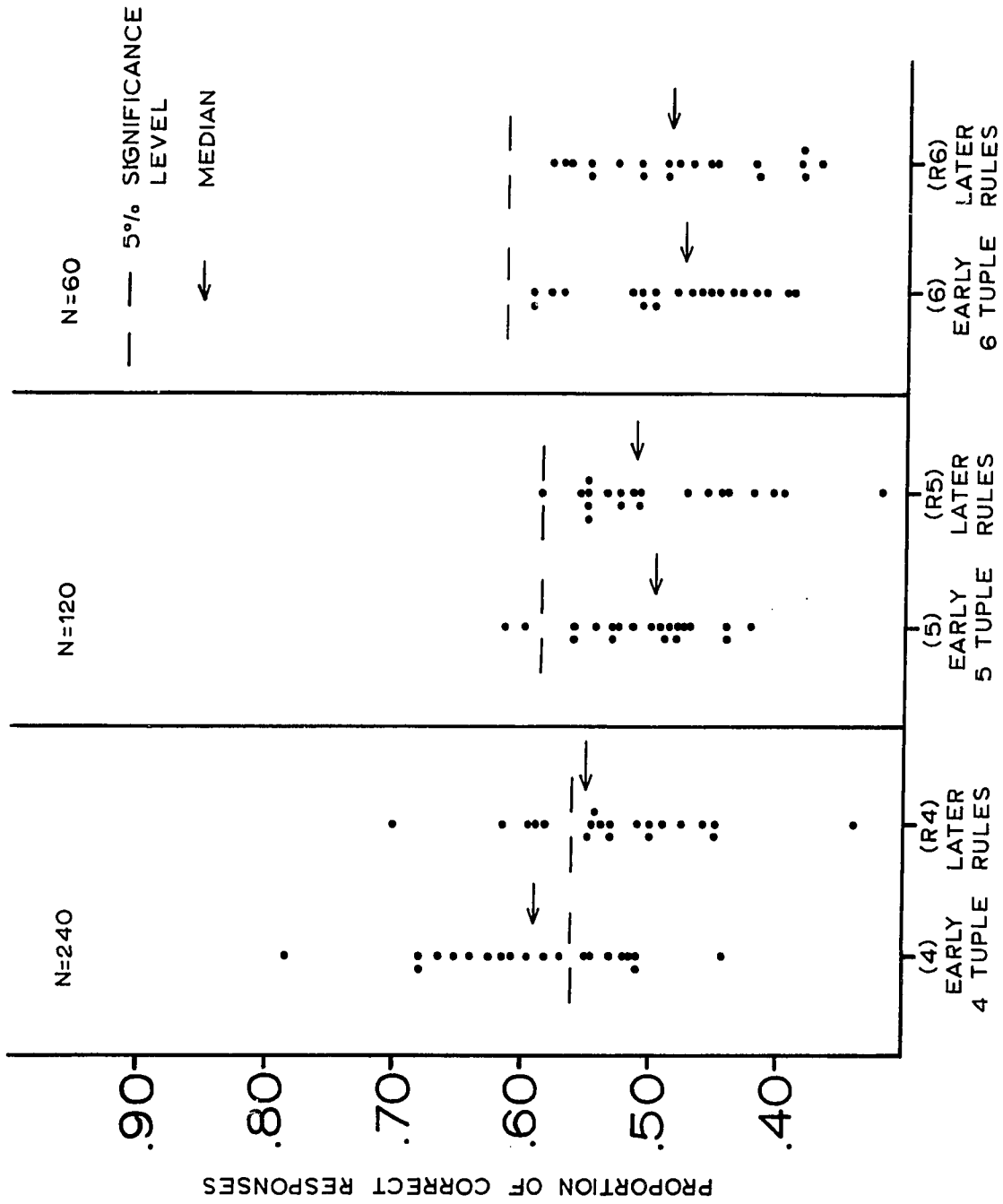


Figure 5. The proportion of correct responses on all k-span rule trials for each subject. Each column presents the data from one of the six experimental groups that had k-span rules in its sequences.

had on k-span rule trials. The proportions of correct responses for the twentieth subject was 0.55 on the run rule trials and 0.68 on the k-span rule trials. Thus for group 4 the run rules were much easier than the k-span rules, in a replication of Rose and Vitz (1966).

A similar comparison of the run rule and k-span rule learning curves for group R4 (see Figs. 2 and 4) showed that the two curves differed significantly only after block 5 when the k-span curve starts to decline, the difference there being only about 0.05. When the overall proportion of correct responses each subject made on the run rule trials was compared with the overall proportion correct he made on the k-span rule trials, of the 20 subjects in group R4, only 9 had a higher proportion of correct responses on run rule trials than they did on k-span rule trials. Although the mean proportion of correct responses on the run rule trials was slightly greater than the mean proportion correct on k-span rule trials, a t test of the difference scores (between the total proportion correct on all run rule trials and the total proportion correct on all k-span rule trials for each subject) was not significant ($t = 1.46$, $df = 19$, $p > 0.1$). Close inspection of the two curves revealed that all eight data points from the run rule trials lie above the corresponding data points from the k-span rule trials. If the two sets of points are from the same populations, then for each pair of points the probability that one point will lie above the other is $\frac{1}{2}$. The probability that all eight data points from one set will lie above the corresponding points in the other set by chance alone is $(\frac{1}{2})^8$ which is less than 0.004. However this effect again is aided by the paradoxical decrease of the learning curve for the k-span rules on the last four blocks of trials. Whether or not this decrease would replicate is problematic. Certainly, if there are any differences between the two curves, they are very small.

No differences between the subjects' performance on the run rule trials and on the k-span rule trials would be expected if the subjects had changed to the k-span code because of the ineffectiveness of the run length code during the random trials. However before any conclusions are drawn about a switch in the coding process, it should be noted that although performance on the run and k-span rule trials was not different, the performance on both rule trials was very low, much below group 4's run rule performance. Therefore the lack of differences may be just a reflection of the general reduction of the subjects' effective memory. Indeed, the orderliness of the run rule data discussed earlier argues for this latter interpretation.

Four of the experimental groups received 502 random trials and the control group received 1506 random trials. The data from these random trials were analyzed in blocks of 251 trials. Data were lost from the control group's last block of trials because one group of five subjects did not finish the sequence in the 100 min. time limit and one other subject left the experiment about 150 trials before the end of the sequence.

Figure 6 presents the recency curves for the first two blocks of trials combined over the four experimental groups and the control group for a total of 100 subjects. The proportion of responses which predicted a run continuation (i.e., response B following run lengths of b and response Y following run lengths of y) are plotted against the run length. Data from the two events were combined. The large number of subjects provides a good estimate of the form of the average recency curve for the first two blocks of trials. The individual points show the performance of each of the five groups. The horizontal bars show the 95% confidence intervals for the combined data.

The recency curve for the first block of trials definitely tends to

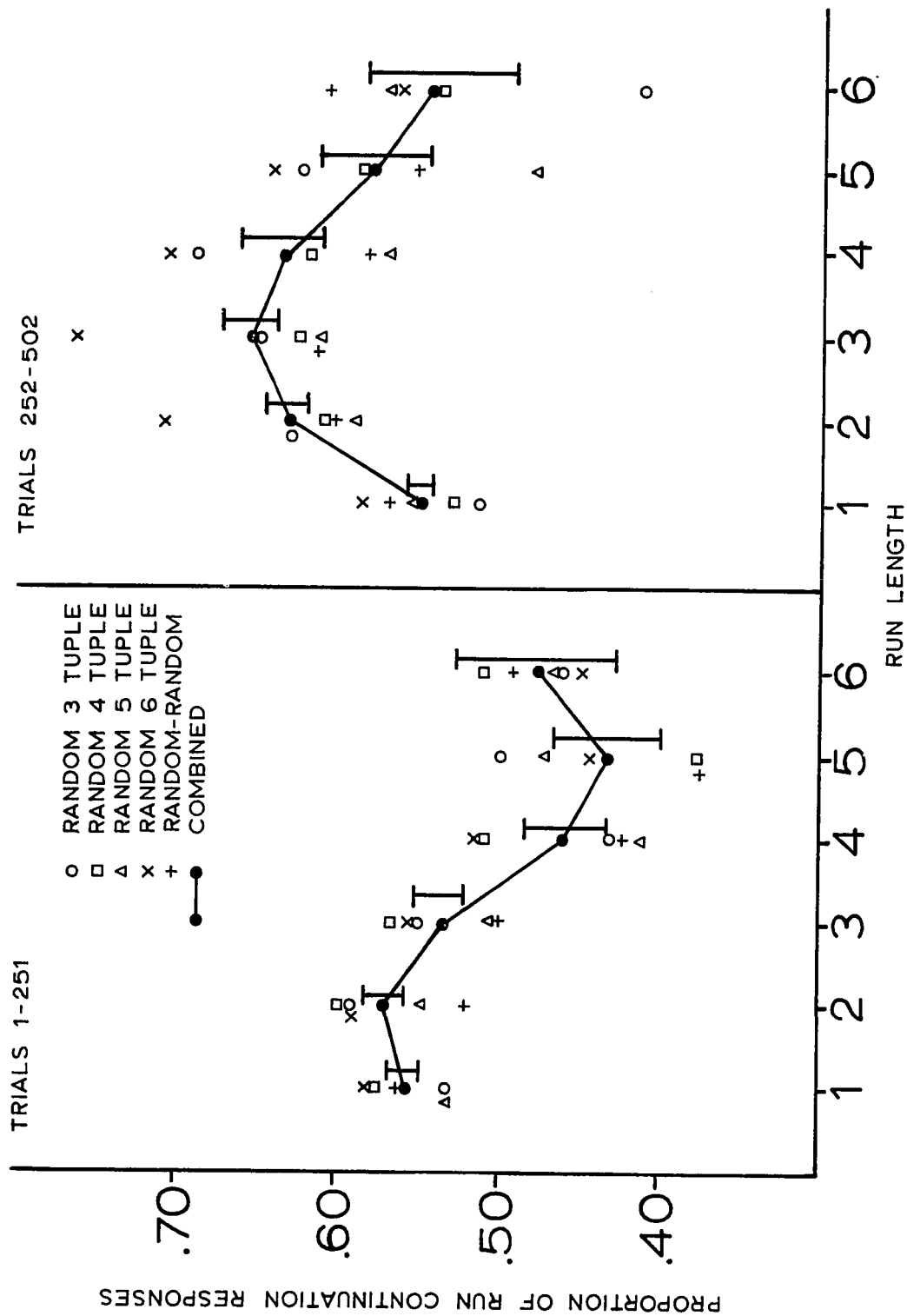


Figure 6. The recency curves for the first two blocks of 251 trials for the five groups which received random sequences on these trials. The solid lines show the combined data from all five groups.

decrease with increasing run length. It decreases at least 0.10 from about 0.56 to 0.46. The second block of trials shows the change over trials taking place. In the second block of trials the proportion of responses predicting a run continuation was significantly greater than the corresponding proportion for the first block of trials after every run length except the first. In addition, the recency curve for the second block of trials has a slope which is positive up to run length three after which it becomes negative. Thus both the occurrence of negative recency effects and the changes toward positive recency effect on later trials are replicated.

Figure 7 shows the control group's recency curves for all six blocks of trials. The wide confidence intervals reflect the reduction in the number of subjects from 100 to 20 subjects, and limit the conclusions that can be drawn. After the first block of trials the recency curves are all fairly flat. The probability of a prediction that the run will continue was approximately 0.55 to 0.60. These data are somewhat different than those reported by Edwards (1961). The slope of the recency curve for the 50:50 sequence in Edwards' experiment was negative on trials 1-200 and flat on trials 201-400, but on trials 401-1000 it showed some evidence of being positive. Derks' (1962) curves for a 75:25 sequence, however, never showed any definite evidence of a positive slope. The 60% repetition level shown in Figure 7 agrees with Anderson's (1960) data. Clearly the negative recency effect was eliminated on the later trials. However, except for the early part of the curve for the combined data in the second block of trials in Figure 6, no positive recency effect was found.

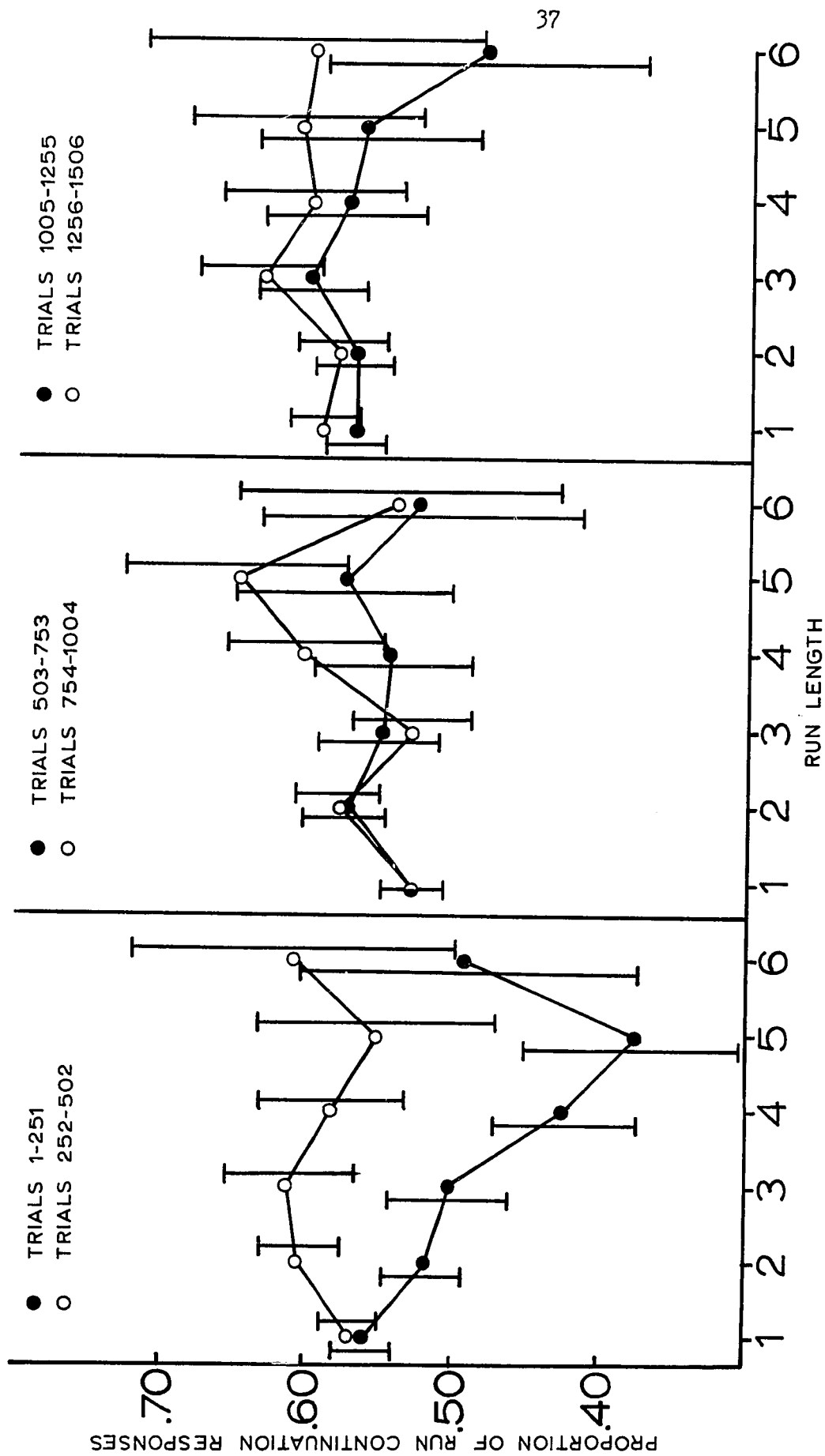


Figure 7. The recency curves for the Random-Random control group in six blocks of 251 trials.

EXPERIMENT II: METHOD

Subjects

The 80 subjects were obtained by advertising on the University of Washington campus. The subjects were paid \$2.00 for the 40 min. session. An extra dollar was offered to anyone who obtained 60% or more correct responses.

Apparatus

A block diagram of the audio circuit is shown in Fig. 8. The output from an oscillator was switched into one of three attenuators. One of the attenuators controlled the intensity of the standard tone; the other two Daven 0.1 dB precision attenuators were set at 1.0 dB above and below the attenuation of the standard intensity tone. A relay with mercury-wetted contacts was used to gate the signal. Switching transients were reduced by passing the signal through a low-pass section of a Spencer-Kennedy variable electronic filter which provided 18 dB per octave attenuation above the cut-off frequency of 1000 Hz. The signal was amplified and sent to three Permaflux PDR-8 earphones connected in parallel. Each earphone was located in the left ear muff of a pair of Wilson Sound Barrier ear muffs.

The testing was conducted in three IAC model 400 sound attenuating booths. Each booth was equipped with a display panel; the face of this panel is shown in Fig. 9. The white light at the top of the panel served as a warning light and came on at the start of each trial. Two green circles (In-Line Read Out display modules) below the warning light marked the observation intervals. Lights number two and three above the two response

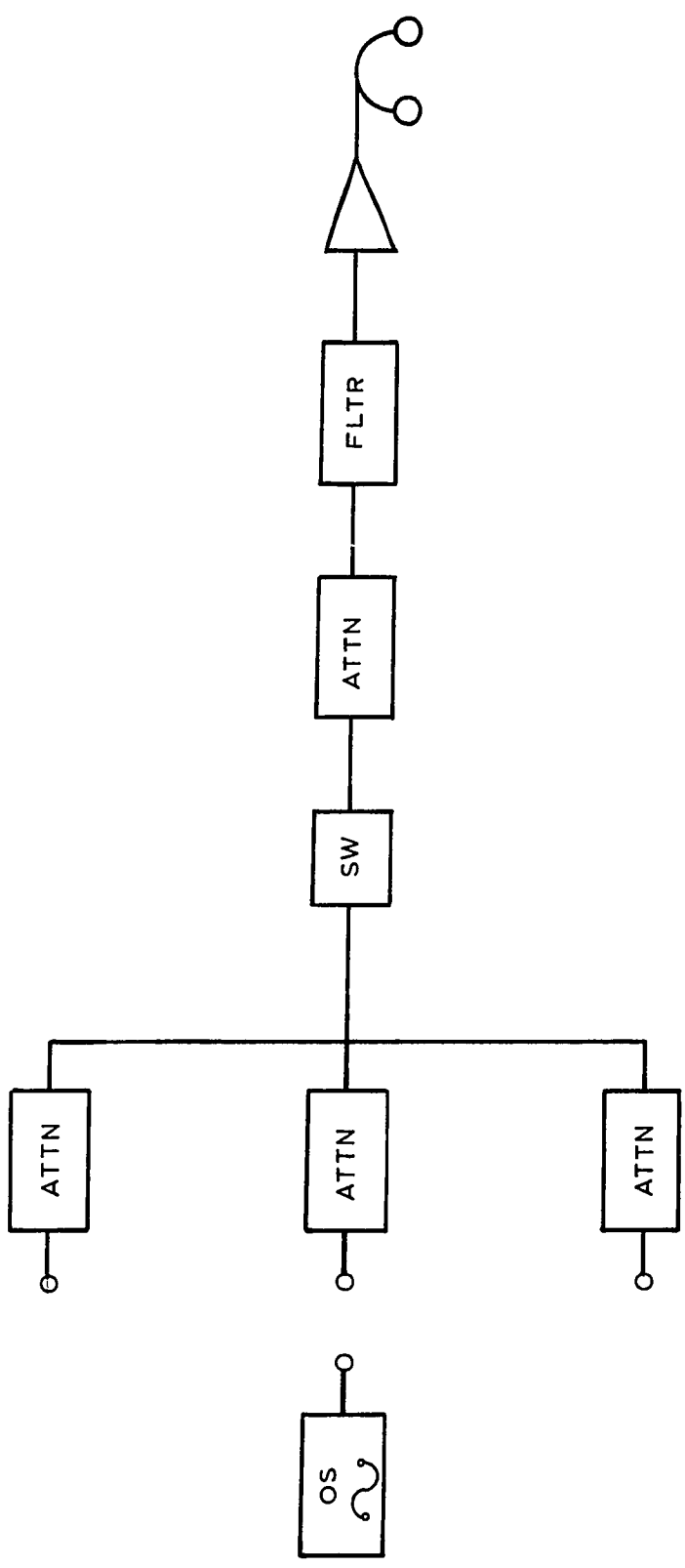


Figure 8. A block diagram of the audio circuit used in experiment II.

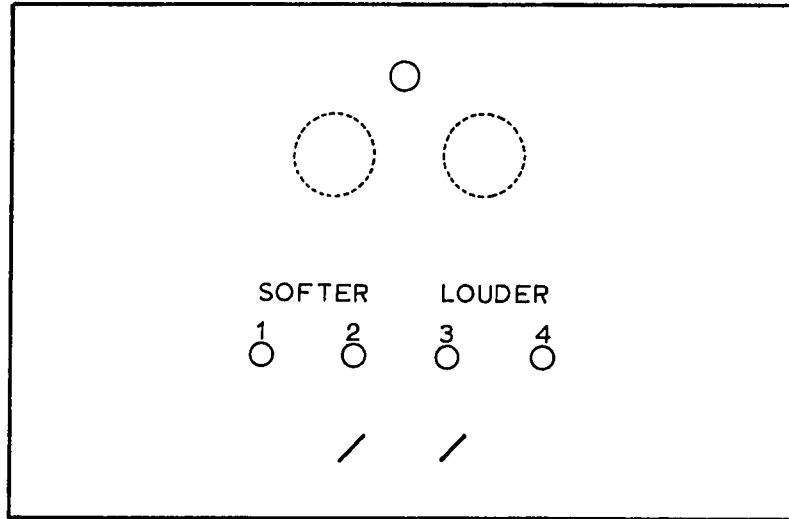


Figure 9. A schematic drawing of the display panel used in experiment II.

keys came on when the corresponding key was pressed during the response interval and were used to inform the subjects that their response had been recorded. The two end lights were used to present the feedback to the subjects. These two lights are considered as events y and b.

Procedure

The subjects were tested in groups of three or fewer. The subjects' task ostensibly was to identify the tone presented on each trial as one that was either more or less intense than the standard intensity tone. They were given the standard intensity tone three times before the experiment started.

Each trial was four sec. long. The warning light came on for $\frac{1}{8}$ sec., and was followed by a $\frac{1}{8}$ sec. observation interval. The subjects had two sec. following the observation interval to make their response. The left key was marked "softer," and the right key was marked "louder." After the response interval, one of the feedback lights came on to inform the subjects whether the tone presented on that trial was louder or softer than the standard intensity tone. The feedback light stayed on for $\frac{1}{8}$ sec. and was followed by a $\frac{1}{2}$ sec. intertrial interval.

The subjects were encouraged several times in the instructions to pay close attention to the signal for the full 40 min. of the experiment, and were offered an incentive to encourage good performance. A one dollar bonus was offered to anyone who could get 60% or more of his responses correct. The subjects were instructed to guess if they were not certain of their choice. They were told that the feedback was there to help them make their decisions. In addition the instructions informed the subjects that the intensities of the tones were difficult to discriminate and that they'

would become more difficult as the experiment progressed.

The standard intensity tone was a 1000 Hz. tone, set at an intensity which was comfortable and clearly audible (approximately 85 dB SPL⁴). For the first 50 trials the subjects received an irregular sequence of 25 trials in which the signal tone was 1.0 dB less intense than the standard and 25 trials in which the signal tone was 1.0 dB more intense than the standard. (The average proportion of correct responses on these trials was 0.70.) The feedback lights were consistent with these differences. For the next 503 trials the standard intensity tone itself was presented on each trial (i.e., there were no stimulus differences) and the sequence of the feedback lights was identical to one of the event sequences of experiment I.

Four groups of 20 subjects were tested. Three of the groups received the first 503 trials from one of the 3, 4, or 5 tuple rule sequences. The fourth group (control) received 503 trials from one of the random sequences. The groups were divided into subgroups and balanced as in experiment I.

Three subjects were discarded and were replaced by three others. One subject was discarded before his data was analyzed because after the experimental session he reported that he did not pay attention at all. Two other subjects were discarded for not making a response on a large proportion of the trials.

EXPERIMENT II: RESULTS AND DISCUSSION

Figure 10 presents both the run rule and the k-span rule learning curves for the three experimental groups. The curves were plotted in the same way as the curves in experiment I were. Again each point is based on 600 observations so that the maximum size of a 95% confidence interval about each point is ± 0.04 . The group curves show little or no evidence

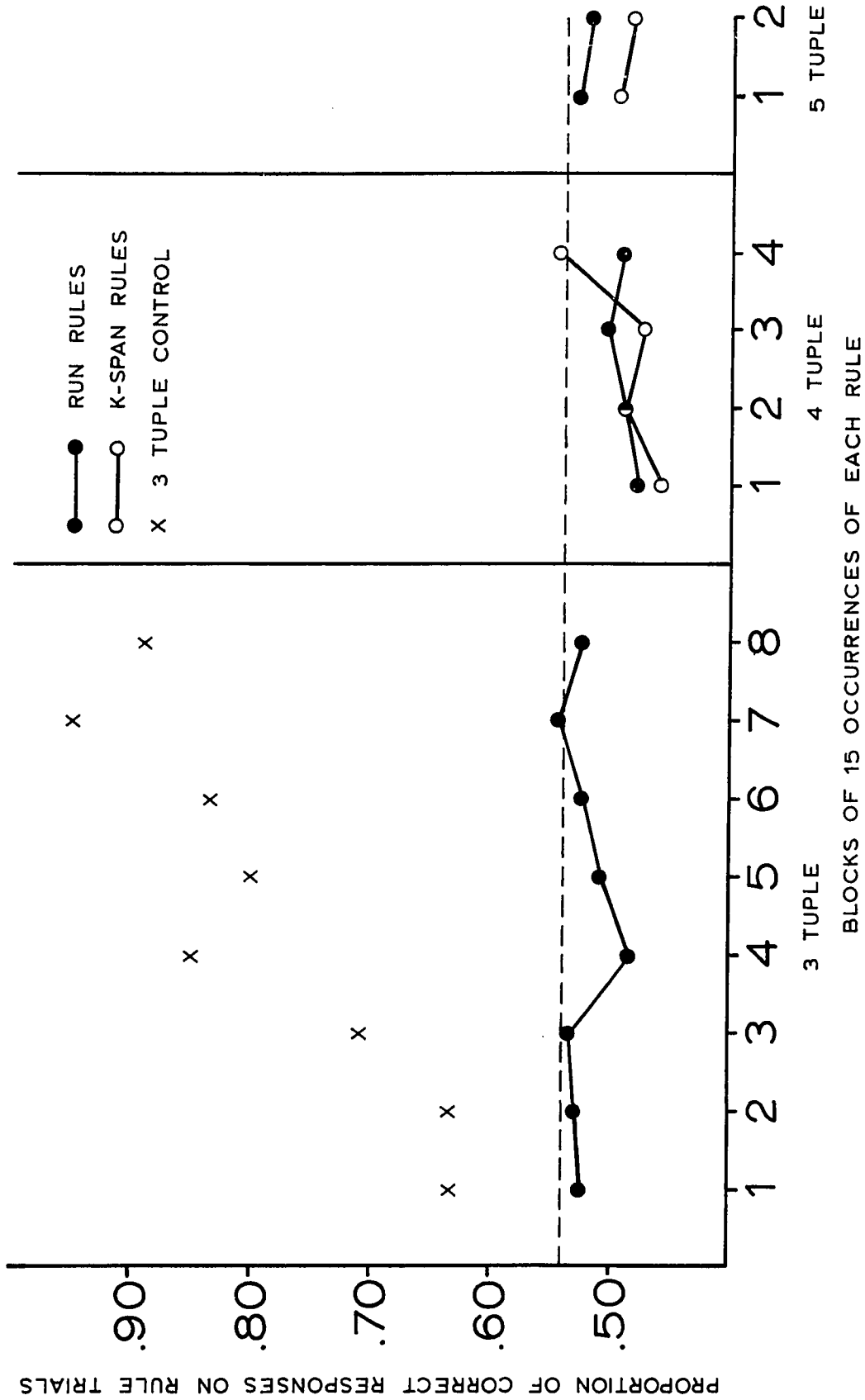


Figure 10. The proportion of correct responses on run rule trials and on k-span rule trials for the three experimental groups. The data points were computed in the same way as those in experiment I were (see Figs. 1, 2, and 4).

of performance above the chance level. Only two data points out of 20 exceed the 95% confidence interval.

Figure 11 presents the proportion correct on all the rules for each subject. The medians and significance levels are shown as in experiment I. Very few subjects show any evidence of performance above the chance level. Only 8 out of 100 subjects performed above the 5% significance level. In addition, one subject in the 3 tuple group whose total proportion correct was only 0.55 showed a rapid improvement on the last two blocks of trials. (23 and 27 correct out of 30) and reported at the end of the session that he found patterns in the sequence. One other subject in the 3 tuple group reported the rules without questioning at the end of the session. His proportion correct was 0.62. The experimenter talked informally about the experiment with many of the subjects at the end of the session. Except for the two subjects mentioned above, none of these subjects seemed to be aware of the rule trials. Thus it is clear that subjects under these recognition task conditions did not learn even the simplest rules used. Apparently, they use very little effective memory in recognition tasks.

One other possibility is that the display and procedure which was used would not lead to learning of the rules even under the best conditions when subjects are instructed to look for the regularities in the sequence. To check on this possibility three subjects were given the 3 tuple sequence under the same conditions as the 3 tuple group subjects, except that they were given no earphones and were instructed to look for regularities in the sequence of event lights. The crosses in Fig. 10 plot the proportion correct obtained by these subjects. Since each data point is based on 90 observations the maximum size of a 95% confidence interval about each point is less than ± 0.10 . Clearly these three subjects performed at a level

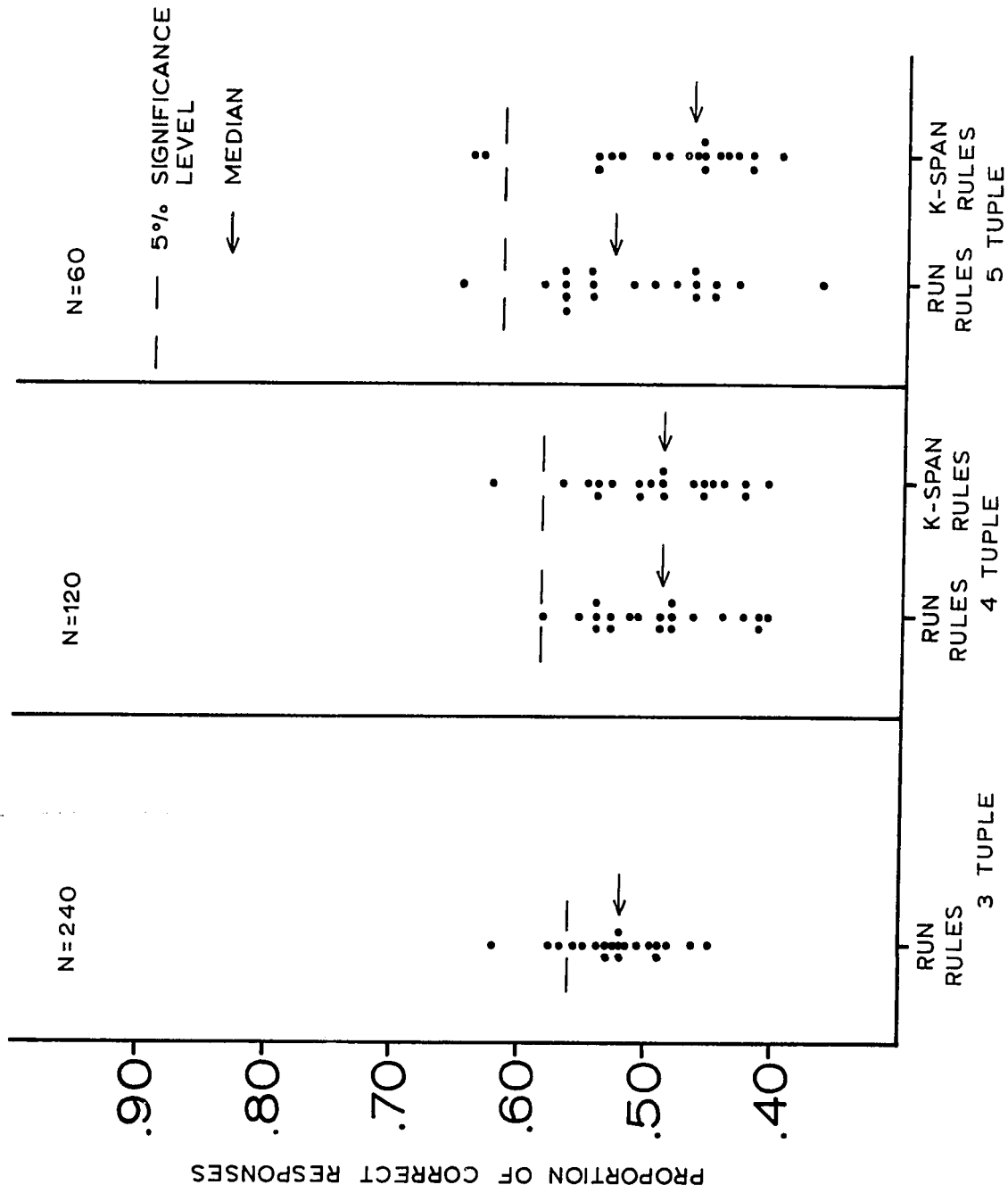


Figure 11. The proportion of correct responses on all run rule trials and, separately, on all k-span rule trials for each subject.

significantly above chance performance.

Figure 12 shows the recency curves for the control group which received the random sequences in blocks of 251 trials. The horizontal bars show the 95% confidence intervals about each point. For comparison the recency curves obtained from the combined data of all the groups on the random trials in experiment I is replotted in the figure. In the first block of trials the control group's recency curve is significantly above the experiment I curve at all points except the first and the last point. The control group's curve rises up to run length three; the drop thereafter is not significant. In the second block of trials the recency curve for the first time appears to have a clearly positive slope. The last two data points are significantly greater than the corresponding data points in the first block of trials. The recency curves changed over trials as the curves obtained in binary choice experiments do.

EXPERIMENT III: METHOD

Subjects

The subjects were 120 undergraduates from the University of Washington. The subjects participated to fulfill part of a course requirement. In addition, \$4.00 prizes were given to the 20 subjects with the best performance.

Apparatus

The five booths from experiment I were used, but they were located in a smaller room. Each booth had two pushbuttons mounted in a small box on the right hand side of the booth's working area. Above each pushbutton was a small red light which came on when the button was pressed during the

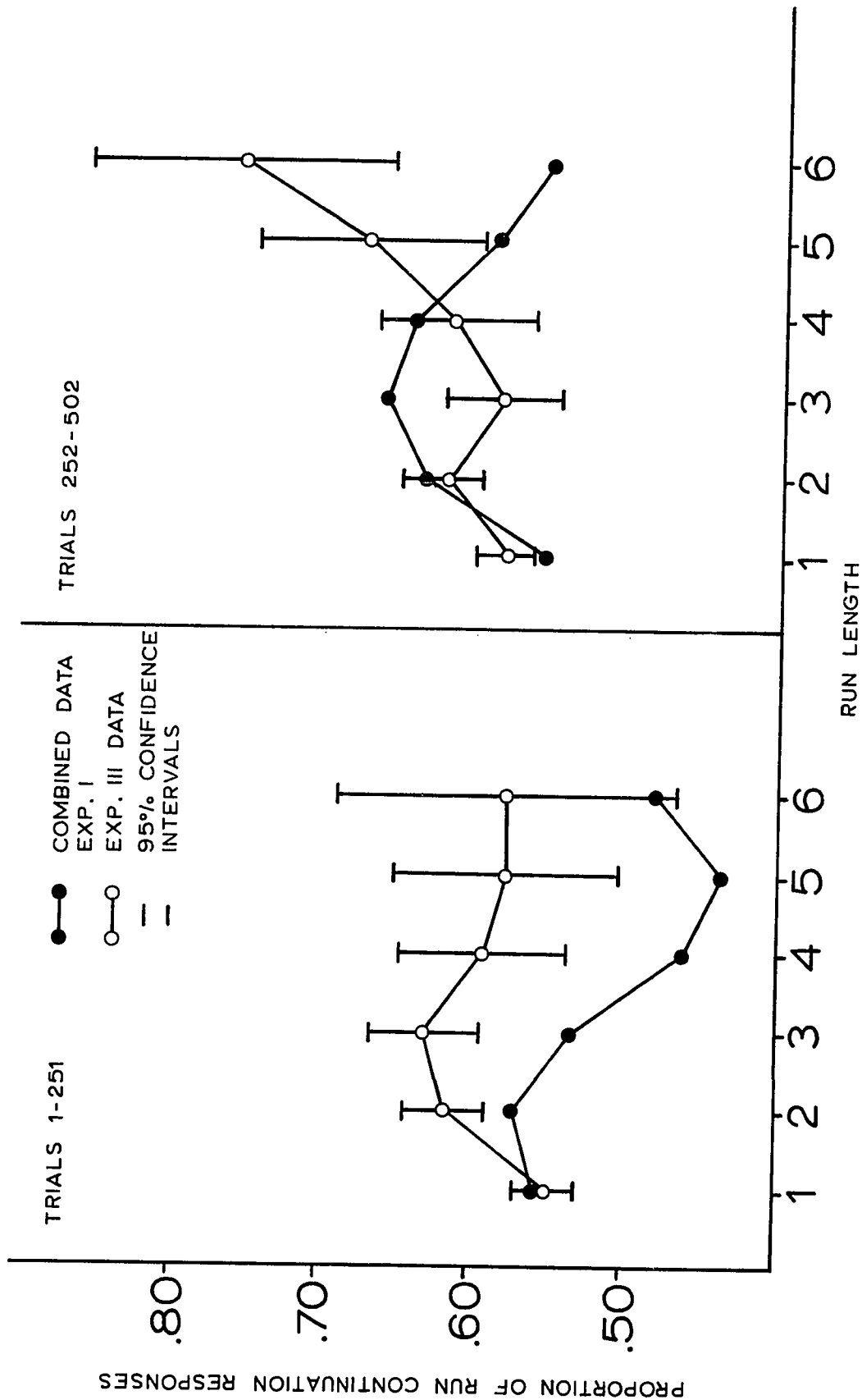


Figure 12. The recency curves for the two blocks of 251 trials for the control group in experiment I. Analogous recency curves from the combined data in experiment I are replotted for comparison.

response interval. In the center of the booth's working area was an answer sheet which contained 15 blank dashes per line. The answer sheet was slid partially beneath a piece of construction paper which was fastened to the table. The answer sheet slid freely back and forth beneath the construction paper and the subjects kept their place on it by covering all the lines which had been already filled in. Also, to help the subjects keep their place they were told when to start a new line on the answer sheet.

A display panel was located about five feet from the subjects. It was the same panel used in experiment I except that some additions were made to it. Figure 13 shows the face of the modified panel. A smaller panel was mounted above the panel used in experiment I. The new panel contained only a single blue light in the center of it. Just below the blue light the words "task 1" were printed. A $\frac{1}{4}$ inch strip of black tape separated the two panels. The words "task 2" were added just below the warning light on the original panel. The control group used only the original panel without any signs. The programming equipment was again located in the same room as the subjects, but was not activated until after the response interval.

Procedure

The subjects were tested in groups of five or fewer. The subjects in the experimental groups were told that the experiment was a test of their ability to do two tasks simultaneously. Task one was a probe memory task similar to the one used by Waugh and Norman (1965). Task two was a binary choice task.

Figure 14 shows the sequence of trial events received by the subjects in the experimental groups. A trial started when the tape recorder presented

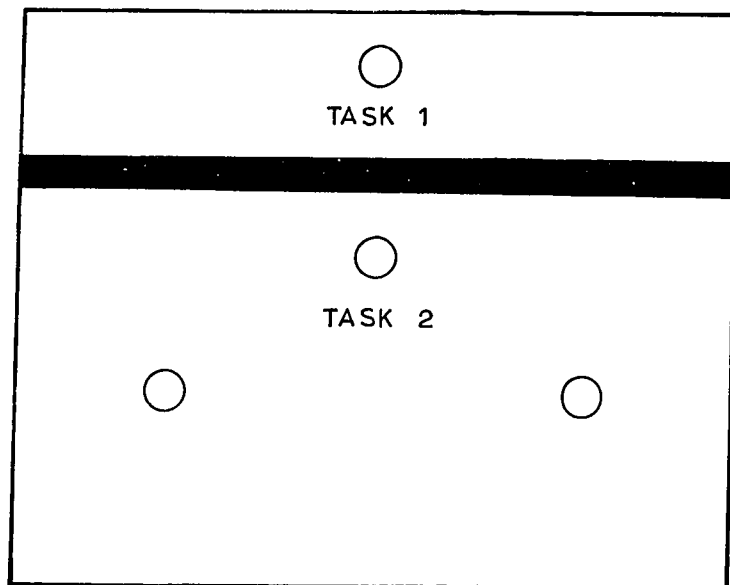


Figure 13. A schematic drawing of the display panel used in experiment III.

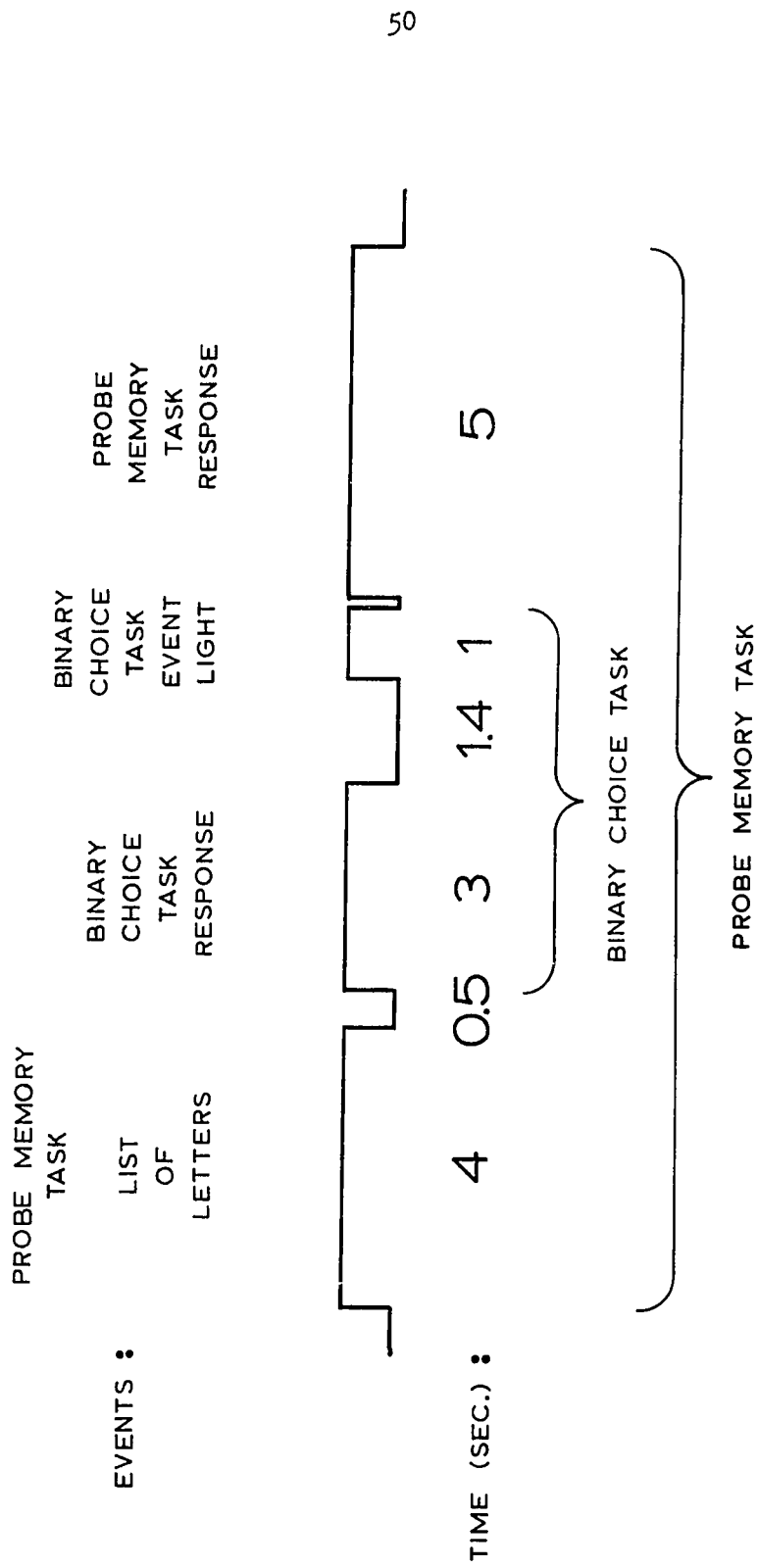


Figure 14. The order of the trial events in experiment III.

a list of letters which the subjects were instructed to remember. The tape recorder was started and stopped by the experimenter who was sitting in front of the room. The list of letters took four sec. to present. About $\frac{1}{5}$ sec. later the yellow warning light labeled task 2 came on and stayed on for three sec. The subjects were to predict which of the two red event lights below the warning light would come on. Only responses made while the warning light was on were recorded. After a 1.4 sec. pause one of the two event lights came on and stayed on for one sec. The blue light labeled task 1 came on next and stayed on for five sec. A single letter from the original list (the probe letter) accompanied the onset of the blue light. The subjects had until the blue light went off to write down the letter which followed the probe letter in the original list. A new list of letters followed the offset of the blue light by approximately 250 msec.

Four groups of 30 subjects were tested. The only difference between the three experimental groups (groups 4, 6, and 8, respectively) was the length of the list of letters they received each trial. The lists were either 4, 6, or 8 letters long. The rate at which the letters were presented varied so that the total presentation time was always four sec. A random number table was used to construct the lists of letters and to choose the positions of the probe letters. The letter w was the only letter of the alphabet which was not used. The probe letter occurred in all positions except the last one. The control group was told nothing about the probe memory experiment, but it had a 9.5 sec. intertrial interval which was the same as the time the experimental groups were given to perform the probe memory task.

All four groups received the same sequences of event lights in the binary choice task. Three different sequences were used. Two of the

sequences were the first 252 trials from the two random sequences used in experiment I; the third sequence was the second 252 trials from one of these random sequences. In each of the three subgroups of each group the left and right events were balanced as in experiment I.

The testing session was divided into three parts. During the first part the subjects were given ten practice trials to acquaint them with the procedure. In the second part the subjects received 134 trials which took about 35 min. It was followed by a five min. rest break during which the subjects were allowed to leave their booths. In the third part the subjects received 118 trials which took another 30 min.

The instructions explained the procedure to the subjects and stressed the independence of the two tasks. The subjects were told to concentrate mainly on the probe memory task and to attempt to do more than guess at the binary choice task only if they could make no further improvements in their probe memory task performance. The subjects were encouraged to make a response to both tasks on each trial.

Prizes were awarded on a basis which was consistent with the instructions. The subjects were told that within each group of 30 subjects they would be ranked according to their performance on the two tasks and the five subjects with the highest rank would receive the prizes. The rank ordering was determined by a lexicographic ordering of the subjects' performance on the two tasks. The scores from the probe memory task were placed into ~~one~~ of ten categories which corresponded to percentiles of the maximum possible score. Within the percentile categories the subjects were ranked according to the number of correct predictions they made on the binary choice task. This ordering was explained to the subjects and several examples illustrating it were given. For the control group prizes were

awarded to the five subjects with the greatest number of correct predictions.

EXPERIMENT III: RESULTS AND DISCUSSION

The recency curves for the groups that received probe memory lists of 4, 6, and 8 letters are shown in Figs. 15, 16, and 17 respectively. The control group's recency curve is shown in Fig. 18. The data from the 30 subjects in each group were combined. The figures show the proportion of left predictions which were made following the run lengths of each event. The proportion of trials on which neither response was made was 0.015, 0.006, 0.020, and 0.008 for groups 4, 6, 8, and the control group, respectively. Since these proportions are small they were ignored. The horizontal bars show the 95% confidence intervals.

The curves for the left and right events were plotted separately because group 8 exhibited a large bias for the left response. The fact that the recency curves for both the left and the right events in Fig. 17 lie above 0.50 reflects this left response bias. The response bias was probably caused by the difficulty of the length 8 list and the asymmetry of the subjects' working area. The response buttons were located on the right side of the booth; the left response button, therefore, was closer and probably easier to press.

The memory change hypothesis predicts that if the reduction in effective temporary memory is just a short term memory effect then a positive recency effect will be found when the subjects' short term memories are reduced sufficiently. No evidence of a positive recency effect can be found in any of the recency curves. For example, it is not possible to draw a positively-sloped straight line that lies within all the confidence limits on any of the curves. In fact, it is not possible on any of the curves

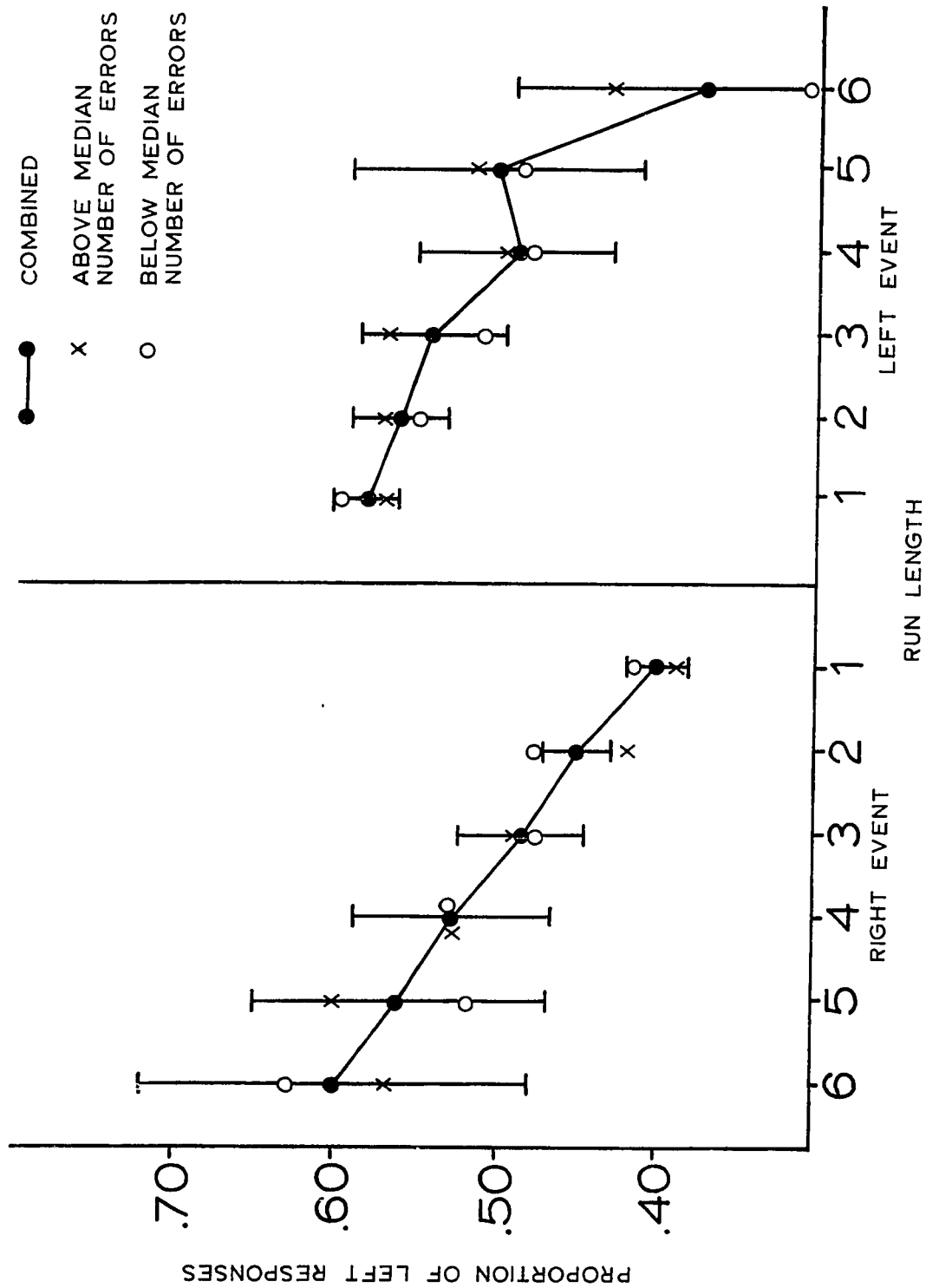


Figure 15. The recency curves for group 4 in experiment III, presented separately for the right and left events.

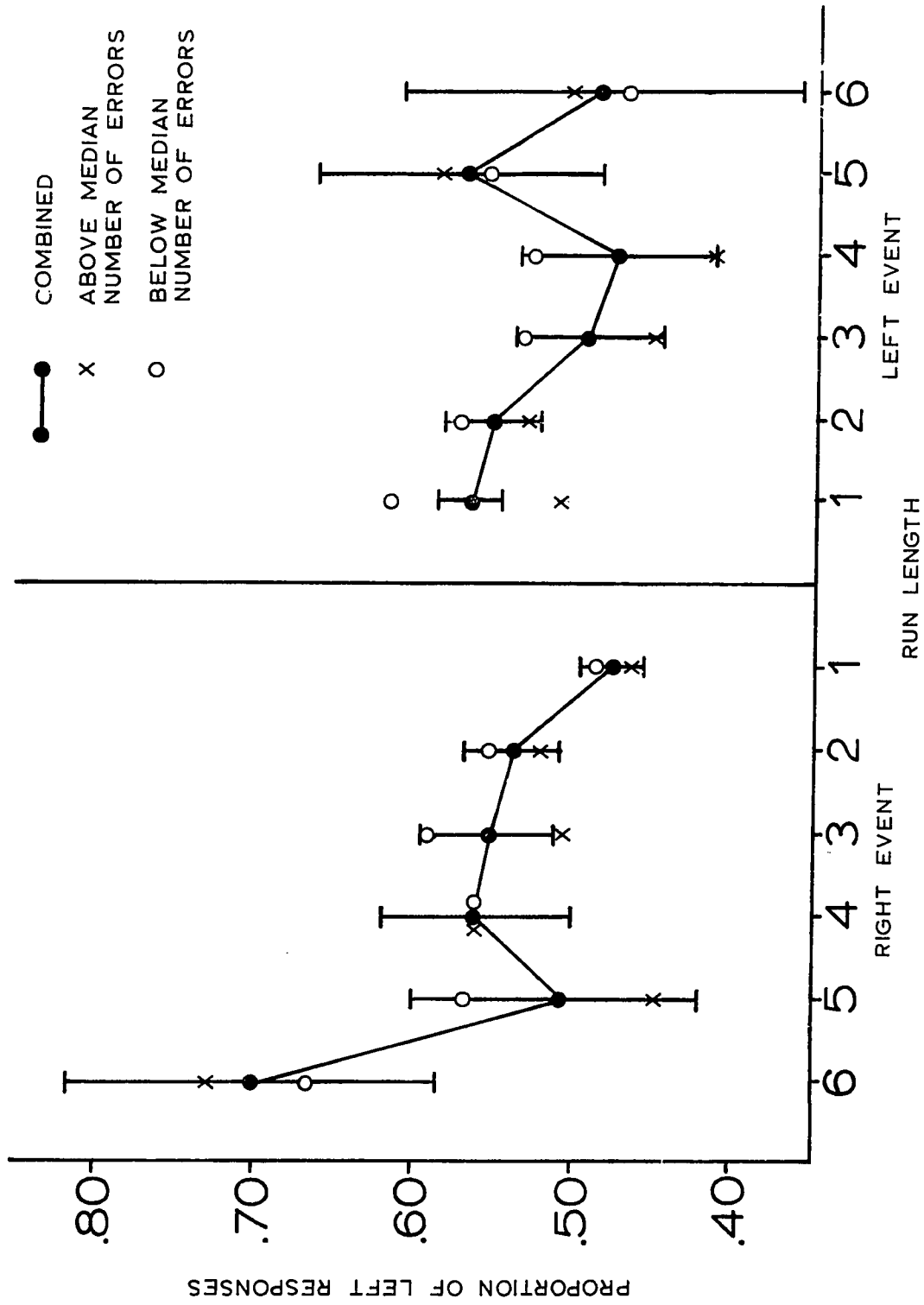


Figure 16. The recency curves for group 6 in experiment III, presented separately for the right and left events.

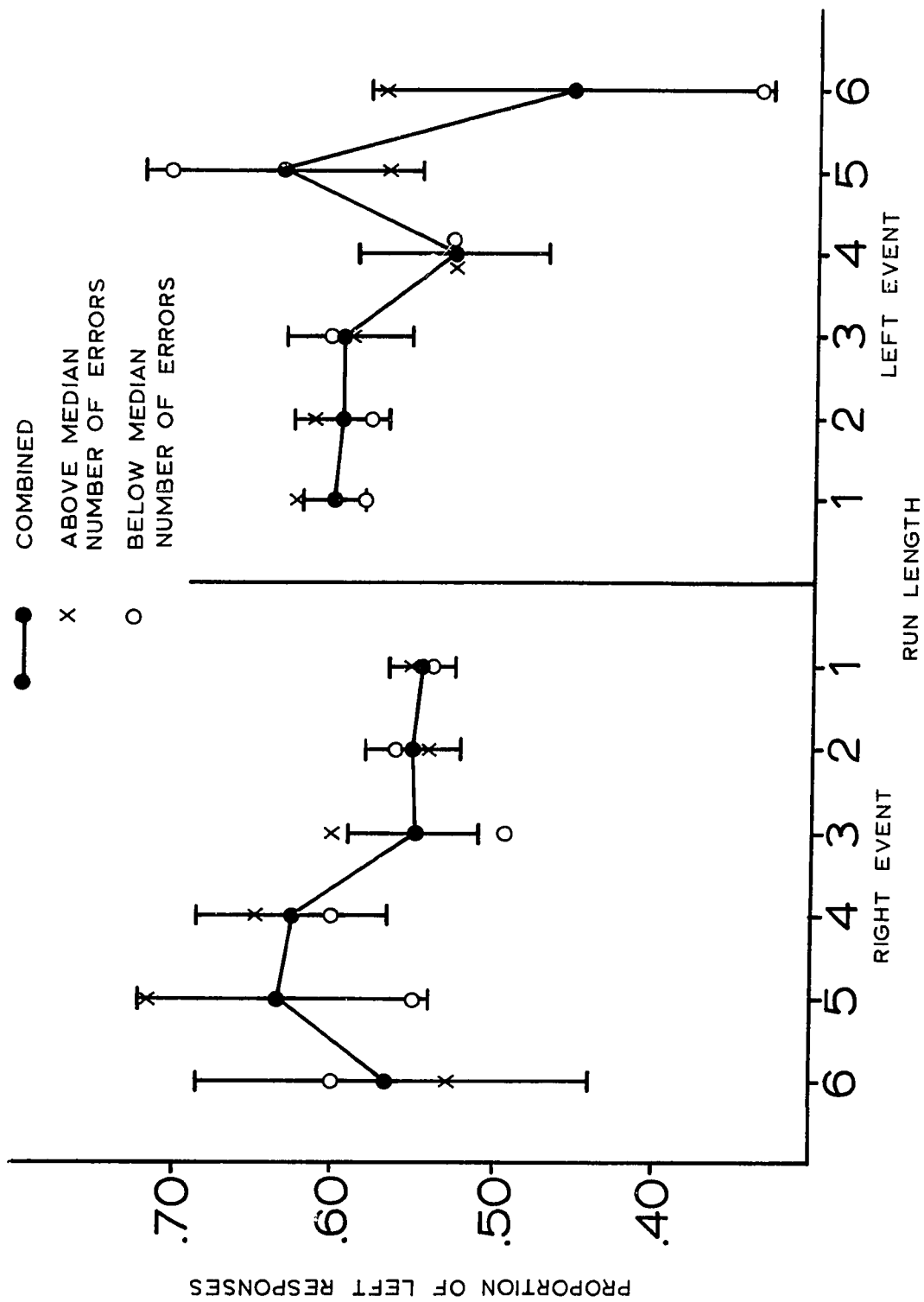


Figure 17. The recency curves for group 8 in experiment III, presented separately for the right and left events.

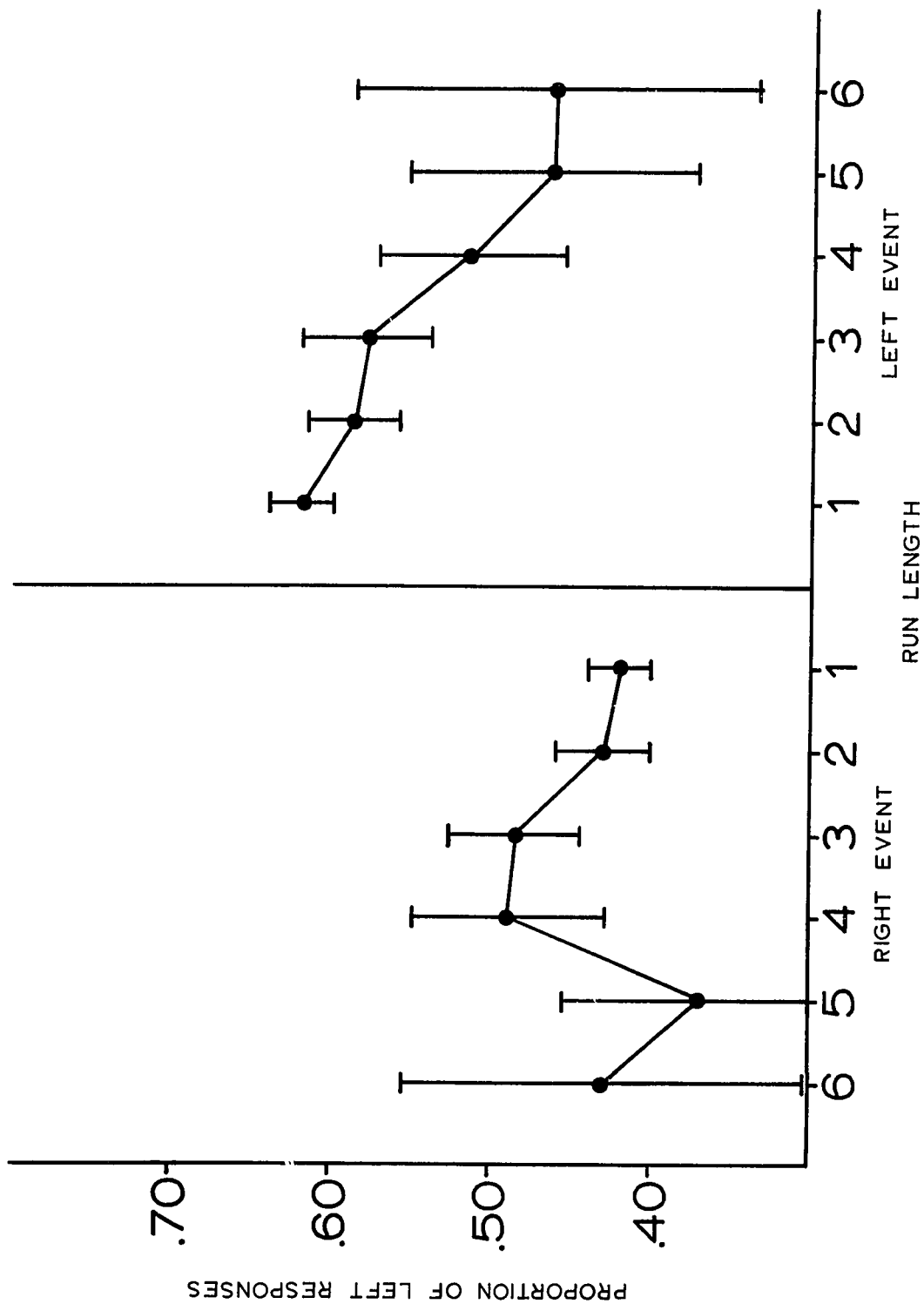


Figure 18. The recency curves for the control group in experiment III, presented separately for the right and left events.

to draw a flat line that lies within all the confidence intervals.

A maximum likelihood estimate of the best fitting straight line was used to obtain a measure of the overall slopes of the recency curves under the assumption that the variance of the response proportions is directly proportional to the run length (Lewis, 1960, p. 181).⁵ Two estimates were made for each of the eight curves. The first estimate used all six data points and the second estimate used only the first four data points. The estimates are presented in Table 2. The negative slopes for the left events and the positive slopes for the right events again provide no evidence for a positive recency effect. The estimates for two of the curves depended on the number of points used. In both of these cases the slope estimated by using all six of the data points was smaller than the slope estimated by using only the first four data points. The differences in both cases were caused by large reversals in both of the last two data points. Only group 8's recency curves appears to have a less negative slope than the others. The estimates of the slopes of the recency curves for group 8 are consistently lower than the estimates from the other groups. Also, notice in Fig. 17 the response proportions for the first three data points in each curve are almost identically equal to each other.

It is possible that the positive recency effect was hidden by individual differences in the subjects' abilities, motivation, or distribution of attention between the two tasks. That is if some of the subjects have a greater memory capacity or greater motivation than the others, the memory task may be less difficult for these subjects and their memory capacities may not be taxed to the same extent as the other subjects. If the recency effects are related to the amount of memory which is used in binary choice tasks then averaging across these two groups of subjects reduces the

Table 2

The estimated straight line slopes for the recency curves obtained from the four groups in experiment III. The slopes were estimated separately for the left and right events and were computed twice, once using all six data points and again using only the first four data points.

Groups	Left	Events	Right	Events
	4 points	6 points	4 points	6 points
Control	-0.030	-0.033	0.026	0.004
4	-0.028	-0.033	0.041	0.040
6	-0.034	-0.015	0.031	0.029
8	-0.019	-0.016	0.019	0.015

expected effect. Another possibility is that some subjects are more likely than others to neglect the probe memory task to some extent and to direct some of their attention to the binary choice task. These subjects would, therefore, show negative recency effects which again, when averaged, would reduce the expected effect. To the extent that subjects neglect the probe memory task their performance on it should suffer. In both of these cases the differences in the subject populations should have been reflected in their performances on the probe memory task. To check on these possibilities the subjects in each group were divided into subgroups on the basis of their probe memory task scores. Subjects who made more than the median number of errors were placed in one subgroup and subjects who made less than the median number of errors were placed in the other subgroup. Recency curves were found separately for the two subgroups of each experimental group and the data points are plotted in Figs. 15, 16, and 17 without any connecting lines. There were no fundamental differences between the two subgroups for any of the experimental groups.

A check was made on the effectiveness of the experimental treatments by comparing the subjects' performance on the probe memory task. The mean number errors made on the 256 trials of the probe memory task was 55.9 for group 4, 122.3 for group 6, and 170.3 for group 8. Orthogonal comparisons of these scores for group 4 vs. group 6, and groups 4 and 6 vs. group 8 were highly significant ($t = 188.5$, $t = 266.4$, respectively; $df = 87$). In addition, at the end of the session the subjects in each experimental group were asked the question: "Did task one exhaust your memory?" Of the 30 subjects in each group the number of subjects replying "yes" was 10 for group 4, 20 for group 6, and 22 for group 8. Thus the experimental treatments were effective in taxing the subjects' memories to different extents. The

poor performance of group 8 makes it unlikely that the recency curves would have become positive with a longer, more difficult list.

To check on the possibility that making a decision on the binary choice task greatly reduced the subjects' performance on the probe memory task, four subjects were tested under the same conditions as the group 4 subjects and five subjects were tested under the same conditions as the group 6 subjects except that they were told nothing about the binary choice task and the warning and event lights were covered. The mean number of errors for the two groups were 48.8 for the lists of 4 letters and 119.2 for the lists of 6 letters. These values are reasonably close to group 4 and 6's means. Therefore, the subjects' decisions on the binary choice task did not severely reduce their probe memory task performance.

FINAL DISCUSSION

On the early trials of binary choice experiments using simple random sequences recency curves have a definite negative slope. When data after trial 500 are analyzed recency curves no longer have a negative slope; the slope is either positive or flat. Preliminary analysis of a recognition experiment conducted by Rose¹ indicated that the slope of the recency curve in this situation was also non-negative. A memory change hypothesis was proposed to account for these differences in the slopes of recency curves. The hypothesis states that the change from a negative to a positive recency effect is caused by a reduction in the subjects' effective temporary memory, the temporary memory being measured with sequence constraints that require a knowledge of the events on the last k trials for solution.

Experiments I and II measured the amount of effective memory subjects

use in the positive and negative recency situations. Both experiments found that subjects' effective memory was reduced in situations in which positive recency effects have been found compared to the subjects' effective memory on the early trials of a binary choice task, where negative recency effects have been found. Thus both experiments provide direct support for the memory change hypothesis.

The experiments also found that although subjects' effective memory was reduced in both situations, the reduction was greater for the recognition experiment than it was for the latter trials of a binary choice task. In the binary choice experiment, after 502 trials on the random sequences, an appreciable number of subjects still learned the 3 and 4 tuple rules. In the recognition experiment, on the other hand, very few subjects learned even the 3 tuple rules. These different estimates of the amount of effective memory being used may, perhaps, be related to the recency curves of the corresponding control groups. In experiment I the recency curves never became positively sloped. After trial 500 the curves were relatively flat at about 0.55 - 0.60. In experiment II, in which a greater memory reduction was found, the recency curve had a slight positive slope on the first 251 trials and an even greater positive slope later. This correlation between memory reduction and the slope of the recency curve suggests that the subjects' effective memory in experiment I was not reduced enough for them to show a definite positive recency effect.

Edwards (1961) reported recency curves for subjects who were tested for 1000 trials on 50:50 sequences. He found that the recency curve was negatively sloped for the first 200 trials and was relatively flat for the next 200 trials. On the last 600 trials the recency curve was positively sloped. In experiment I, on the other hand, in 1500 trials the control

groups' recency curves were still flat and they did not appear to be rising. It is not clear why Edwards' curves became positively sloped and the curves in experiment I did not. There are several differences in the procedures used in the two experiments, but none of them can be related easily to the differences in the results. One of the major differences between the two experiments was in the difficulty of making a response. In Edwards' experiment on each trial the subjects had to lift two corks and make a mark on IBM mark sense answer sheets. As a result it took about three hours for the 1000 trials. In experiment I the subjects merely had to press a key which was in front of them and the majority of the groups finished 1506 trials in 100 min. Two other possible explanations for the different results are: (a) there is a trials by time interaction and (b) the mechanics of making a response influences the results. The mechanics of making a response might influence the results in the following way. In experiment I the keys were always before the subjects and a subject could repeat the response he made on the last trial simply by holding the key down until the warning light came on again. The fact that it was slightly easier to repeat a response may have led to some inertia in the sequence of responses such that the subjects tended to repeat the response made on the last trial. If this inertial tendency was independent of the event sequence, it would lead to a reduction in the slope of the recency curve, making the curve look more flat. In Edwards' experiment responses could not be repeated so easily and his data may be free from this confounding influence.

In experiment II, the recognition experiment, the slope of the recency curve appeared to become somewhat larger on the second block of 251 trials. If the memory change hypothesis is correct this change in slope should have been caused by a further reduction in the subjects' effective memory below

the level used in the first block of trials. Since the 3 tuple rules were introduced in the first block of trials and were not learned, the reduction could only have taken place below the level of effective memory measured by the 3 tuple rules. To solve a 3 tuple rule subjects have to take into account the events which occurred on at least the last three trials. Since the subjects in experiment II did not solve the 3 tuple rules, they could have been utilizing at most the events on the last two trials. Thus if the memory change hypothesis is correct then the increased slope found in the second block of trials in experiment II was due either to a reduction from a memory for the events on the last two trials to a memory for only the event on the last trial (or a further reduction), or to a reduction from a memory for the event on the last trial to no effective temporary memory at all.

When subjects have no effective temporary memory it means that they do not use any of the past sequence of outcome events as discriminative stimuli. This "no memory" state is a necessary condition if the simple path independent models are to apply to binary choice data. Thus, the results of experiment II and my interpretation of them suggest that the situations in which positive recency is observed may perhaps be described by a path independent model. On the other hand, the mere absence of a negative recency effect is no guarantee that one of these models applies. For example, in experiment I no evidence of a negative recency effect was found after the first two blocks of trials. Yet, when rules were introduced at this point they were quickly learned by a considerable proportion of the subjects.

There exists evidence that the path independent models may not be appropriate descriptions of subjects' learning processes in psychophysical

experiments. Friedman and Carterette (1964) found that subjects in a two-alternative forced choice detection experiment were sensitive to first order conditional probabilities in the sequence of outcome events. In their experiment the probability of event repetition or even alternation was varied while the overall event probability was held constant. The influence of these manipulations could be seen in the subjects' recency curves. Thus, based on the limited amount of evidence available, subjects in psychophysical experiments have an effective memory for the events on at least the immediately preceding trial but not for events more than two trials back.

That subjects have a memory for the event which occurred on the immediately preceding trial is not unreasonable, given the ease with which a binary event may be coded and stored. Any slight change in the pressure of the fingers on the response keys, in the tilt of the head, or in the direction of the subjects' gaze may not only be stored easily for one trial, but it also has a straightforward relationship to one of the outcome events. A temporary memory for the event which occurred on the last trial may be a rather effortless achievement.

It would be interesting to determine whether subjects have an effective memory for the event which occurred on the last trial when they are deprived of these position cues. Depriving subjects of position cues only removes the straightforward relationship between the physical changes the subjects may be using as a one trial binary memory and the outcome events, i.e., the compatibility of the outcome events with the physical changes which are responsible for the one trial binary memory is reduced; it does not impair the subjects' ability to use these physical changes as a one trial binary memory. However, the subjects' problem lies in

interpreting this binary memory. There is something compelling about associating, for example, a left tilt of the head with a left event and a right tilt with a right event. When position cues are eliminated the outcome events are defined in a way which makes them independent of position. Tilts of the head may occur just as easily as before but before they can be utilized each direction of tilt must be associated with one of the outcome events. If subjects form artificial associations between these physical changes and the outcome events or if subjects' memory for the immediately preceding outcome event is verbal, then depriving subjects of position cues will have no influence on their effective memory for the immediately preceding outcome event.

Position cues can be eliminated from the response by varying the label of the response keys from trial to trial; they can be eliminated from the outcome events by using a single display which can be in one of two qualitative states. For example, on each trial the two response keys could be labeled "L" and "S" for loud and soft, respectively. The "L" and "S" would switch positions from trial to trial. The outcome event could be an "L" or an "S" presented in a single center display panel. A somewhat less confusing way to eliminate position cues from the subjects' responses is to have them report their choices verbally. It is possible that without position cues subjects have no effective temporary memory and that one of the path independent models would be applicable.

There are, at present, no data on whether subjects in psychophysical experiments have an effective memory for the two immediately preceding events. The techniques used in experiments I and II to measure the subjects' effective memory cannot be applied because the sequence degenerates into a double alternation sequence. Although a double alternation sequence

technically requires a memory for the last two events to be solved, it somehow seems to be different from the tuple rules which do not apply on every trial. Two measurement techniques which could be used are: (a) to put run rules in only one of the two events and (b) to vary the second order conditional probabilities in the sequence of outcome events. Note that these two measurement techniques are not completely equivalent measures of the subjects' effective memory. Data on 2 tuple rules would be useful for evaluating models which assume that subjects use only the events on the immediately preceding trial (e.g., see Tanner, Haller, and Atkinson, 1967).

In experiment III an attempt was made to change the slope of the recency curve on the early trials of a binary choice task from negative to positive by varying the difficulty of an irrelevant short term memory task. No support for the memory change hypothesis was found, although the subjects' verbal reports, their performance on the memory task, and the presence of a large response bias in the most difficult condition all indicated that the experimental treatment was effective in reducing the amount of short term memory capacity the subjects had available for use. One of the main limitations of the study was the complexity of the instructions. It is not clear whether the subjects could divide their attention as the instructions asked and maintain this division for the entire session.

On the other hand, I will show that a possible mechanism for subjects' effective temporary memory has a property that may have caused the failure of experiment III to support the memory change hypothesis. When experiment III was designed the concept of effective temporary memory was developed no further than it was developed for experiments I and II; it was considered as an empirical measure of the subjects' ability of learn rule tuples of different lengths, this ability being dependent on the subjects' short term

memory system. A possible mechanism will now be considered.

Earlier, for explanatory reasons, a description of a possible mechanism for the subjects' effective temporary memory was given. The mechanism was simply that subjects count the number of times each event occurs successively, resetting their count whenever a new run begins. The count on the previous trial must be remembered temporarily in a short term store and either advanced or started anew depending on whether the run of events continues or ends. Evidence that subjects try to solve sequences by counting was cited earlier (Warden, 1924). In addition, there are several informal reports in the literature that subjects spontaneously emit phrases such as "I must have miscounted" (Rose and Vitz, 1966; Vitz and Todd, 1967).

If subjects are counting, then they should be making the same kinds of errors that people make when they tally items. For example, if a psychologist wants to find out the number of Y responses his subjects made, he sits down with the response protocol in front of him and moves his finger through the list of items until he comes to a Y response. Each time his finger reaches a new Y response he advances his count by one. This goes on until he finishes counting the responses, or until he makes a noticeable mistake. Casual observation suggests that when people tally events mistakes are usually made by "losing your place." That is you may have a count in mind and may also have your finger on one of the events to be tallied, but you may not be sure of whether your finger and the count correspond. "Forgetting" seems to occur not by forgetting what the count was but by misplacing the count forwards or backwards by one. Questions such as, "Did I count and not move my finger" and "did I Move my finger and not count" may arise.

In the binary choice task, however, instead of the subjects having control over the sequence of successive items the experimenter controls it, i.e., the experimenter "moves their fingers for them" Therefore they never "forget to move their fingers." However the subjects may have lapses of attention and forget to advance the count. If subjects are unaware of their lapses or if they make no attempt to correct for a possible error, then a subject's count will always be less than or equal to the actual count. Thus if a subject makes an error in counting a run of events, then, depending on the sequence being used, the error may first be observed in his behavior on some later trials. For example, if subjects are presented only runs of length one and four randomly interspersed (sequences like ybyyyybybbby . . .) then it is possible to predict correctly on every trial except those following runs of length one. However, if a subject has learned the pattern of events but makes a counting error, his error, even if it occurred on one of the previous trials, will first show up in his behavior on the trial following the end of the ongoing run length. On this trial, although the actual count is four and the appropriate rule is a count of four terminates, if a subject has made one counting error his count is three and the rule that would be applied is a count of three continues. Prior to the last trial an error in his count led to the same response which would have been made had the correct rule been applied. Hence an error in counting could show up one, two or three trials later than it was actually made. This example also illustrates another property of this proposed forgetting mechanism. The subject in the above example observed, according to his count, that a count of three terminated; thus, there appeared to be a rule violation. Such subject-induced rule violations would decrease correct performance on trials prior to the end of

the ongoing run length.

For the purpose of demonstrating one of the properties of this proposed forgetting mechanism, assume that a subject has learned the rules and is not influenced by his self-induced rule violations. If the subject can be characterized by a parameter, c , which is the probability that on any trial he will forget to advance his count, and if run terminations are especially noticeable so that the subject always records the start of a new run, then the probability that he will make an incorrect response following run length k , where k is the longest run length in the sequence, is $1 - (1-c)^{k-1}$, the probability of making at least one counting error on the previous $k-1$ trials, a geometric increase. In real data the exact form of the increase will depend, of course, on the details of the learning process which describes how the counts are stored and on the sequence of events which is used, but the rapid geometric increase will have a strong influence on the data.

Restle (1966) tested subjects on sequences of events which contained only two different lengths like the sequence described above. He presented data on subjects' performance on pairs of runs of various lengths and his data show a sharp increase in the proportion of incorrect responses on the trial following the longest run length, as the length of the longest run increased (0.09 for $k = 3$, 0.27 for $k = 4$, 0.35 for $k = 5$, 0.43 for $k = 8$, and 0.42 for $k = 0$). The other rule trials did not show this sharp increase. Restle's data provide at least qualitative support for the proposed forgetting mechanism. This forgetting mechanism will not be developed further in the present paper. There are many possible modifications of it that can be tested more directly on sequences which are simpler than those used in the present experiments.

In terms of the proposed forgetting mechanism, the memory change hypothesis states that when subjects are tested for many trials the parameter c increases, i.e., they have an increased probability of forgetting to advance their count. This increased forgetting reduces subjects performance on rule trials. In particular, if the parameter is increased sufficiently before the rules are introduced then the rules are unlikely to be detected because of the large number of incorrect counts. Recency curves presumably change both because of the direct influence of miscounting and also because the subject-induced changes in the sequence of counts are treated as changes in the sequence.

If subjects' effective temporary memory is accomplished by counting successive events then the absence of a positive recency effect in experiment III is less surprising. Subjects can remember the length of an entire run of events simply by remembering one number. Since this number holds so much potential information it may be given priority in short term memory over one of the letters from the memory task. The probe memory task which was used probably never eliminated this simple count. A more adequate task may be one in which the subject is continuously occupied with verbal material which has to be processed at a rapid rate, a task like the shadowing task (Cherry, 1953).

Thus experiment III suggests that the subjects' effective temporary memory is not just a short term memory phenomenon. To account for experiment III something like the counting mechanism which was proposed must be brought in. The literature to date has not stressed the possibility that subjects are counting. However this remains a strong possibility that should certainly be investigated more fully and, if substantiated, should be given more explicit recognition.

BIBLIOGRAPHY

- Anderson, N. H., Effect of first-order conditional probability in a two-choice learning situation. J. exp. Psychol., 1960, 59, 73-79.
- Atkinson, R. C. and Estes, W. K. Stimulus sampling theory. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), Handbook of Mathematical Psychology, II. New York: Wiley, 1963, pp. 121-268.
- Brown, W. L., and Overall, J. E. Implications of recency effects for probability learning theories. J. gen. Psychol., 1959, 61, 243-251.
- Burke, C. J. and Estes, W. K. A component model for stimulus variables in discrimination learning. Psychometrika, 1957, 22, 133-145.
- Bush, R. R., and Morlock, H. C. Test of a general conditioning axiom for human two-choice experiments. Memorandum MP-1. Department of Psychology, University of Pennsylvania, 1959.
- Bush, R. R., and Mosteller, F. Stochastic Models for Learning. New York: Wiley, 1955.
- Chapman, Jean. The spacing of sequentially dependent trials in probability learning. J. exp. Psychol., 1961, 62, 545-551.
- Cherry, E. C. Some experiments on the recognition of speech, with one and two ears. J. acous. Soc. Amer., 1953, 25, 975-979.
- Coombs, C. H. A portfolio theory of risk preference. Mich. Math. Psychol. Prgrm. Tech. Report MMPP 68-5, 1968.
- Derks, P. L. The generality of the "conditioning axiom" in human binary prediction. J. exp. Psychol., 1962, 63, 538-545.
- Derks, P. L. Effect of run length on the "gambler's fallacy." J. exp. Psychol., 1963, 65, 213-214.
- Edwards, W. The theory of decision making. Psychol. Bull., 1954, 51, 380-417.

- Edwards, W. The prediction of decisions among bets. J. exp. Psychol., 1955, 50, 201-214.
- Edwards, W. Probability learning in 1000 trials. J. exp. Psychol., 1961, 62, 385-394.
- Engler, Jean. Marginal and conditional stimulus and response probabilities in verbal conditioning. J. exp. Psychol., 1958, 55, 303-317.
- Epstein, R. A. The Theory of Gambling and Statistical Logic. New York: Academic Press, 1967.
- Estes, W. K. Of models and men. Amer. Psychologist, 1957, 12, 609-617.
- Estes, W. K. Learning theory. Ann. Rev. Psychol., 1962, 13, 107-144.
- Estes, W. K. Probability learning. In A. W. Melton (Ed.), Categories of Human Learning. New York: Academic Press, 1964, pp. 89-128.
- Feldman, J. On the negative recent hypothesis in the prediction of a series of binary symbols. Amer. J. Psychol., 1959, 72, 597-599.
- 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, Calif.: Stanford University Press, 1965, pp. 250-316.
- Friedman, M. P., and Carterette, E. C. Detection of Markovian sequences of signals. J. acous. Soc. Amer., 1964, 36, 2334-2339.
- Gambino, Blase, and Myers, J. L. Effect of mean and variability of run length on two-choice learning. J. exp. Psychol., 1966, 72, 904-908.
- Goodnow, Jacqueline J. Response sequences in a pair of two-choice probability situations. Amer. J. Psychol., 1955, 58, 624-630.
- 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.

- Jarvik, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. J. exp. Psychol., 1951, 41, 291-297.
- Jones, M. R., and Myers, J. L. A comparison of two methods of event randomization in probability learning. J. exp. Psychol., 1966, 72, 909-911.
- Keller, L. Run structure and the learning of periodic sequences. Unpublished doctoral dissertation, Indiana University, 1963.
- Lewis, D. Quantitative Methods in Psychology. New York: McGraw-Hill, 1960.
- Lindman, H., and Edwards, W. Supplementary report: Unlearning the gambler's fallacy. J. exp. Psychol., 1961, 62, 630.
- Nicks, D. C. Prediction of sequential two-choice decisions from run events. J. exp. Psychol., 1958, 57, 105-114.
- Restle, F. The Psychology of Judgment and Choice. New York: Wiley, 1961.
- Restle, F. Run structure and probability learning: Disproof of Restle's model. J. exp. Psychol., 1966, 72, 382-389.
- Restle, F. Grammatical analysis of the prediction of binary events. J. verbl. learn. verbl. Behavior, 1967, 6, 17-25.
- Rose, R., and Vitz, P. The role of runs in probability learning. J. exp. Psychol., 1966, 72, 751-760.
- Rubenstein, I. Supplementary report: The influence of one stimulus on the prediction of the alternative stimulus in two-choice problems. J. exp. Psychol., 1961, 62, 311-312.
- Tanner, T. A., Haller, R. W. and Atkinson, R. C. Signal recognition as influenced by presentation schedules. Percept. and Psychophysics, 1967, 2, 349-358.

Vitz, P. C., and Todd, T. C. A model of learning for simple repeating binary patterns. J. exp. Psychol., 1967, 75, 108-117.

Warden, C. J. The relative economy of various modes of attack in the mastery of a stylus maze. J. exp. Psychol., 1924, 7, 243-275.

Waugh, N. C., and Norman, D. A. Primary Memory. Psychol. Rev., 1965, 72, 89-104.

FOOTNOTES

¹This study is still unpublished. Personal Communication. Richard M. Rose, University of Washington, Seattle, Washington.

²Although there are 2^n different n-tuples, because of the symmetry of the outcome events $\frac{1}{2}$ of these tuples are equivalent to other tuples. In addition, in each sequence there are two run rules. Hence there are $2^{n-1} - 2$ possible k-span rules.

³The proportions were obtained by using 30 responses from each subject. Since the responses may not have been independently generated, the independence assumption of the De Moivre - La Place theorem may not have been satisfied and the estimate of the significance of the proportions which was obtained from its use may be in error. However for the majority of the comparisons the differences were either very small or so large that they were significant when evaluated by Chebyshev's Inequality.

⁴The outputs of the three earphones were within 4 dB of each other.

⁵When the variance is constant, the maximum likelihood estimate is the same as the familiar least squares estimate. Under the assumption that the variance is proportional to the abscissa values, each squared difference is weighted by the reciprocal of the abscissa value. The assumption that the variance is directly proportional to the run length was made because the number of observations is inversely proportional to the run length and the variance is inversely proportional to the number of observations. The computational formula was

$$m = \frac{\sum Y_i \sum \frac{1}{X_i} - N \sum \frac{Y_i}{X_i}}{\sum X_i \sum \frac{1}{X_i} - N^2} ,$$

where X_i are the abscissa values, Y_i are the values of the ordinate, and N is the number of observations.

VITA

Herbert Anthony Colle was born in Milwaukee, Wisconsin on August 27, 1943 to J. Edgar and Martha Colle. He received his secondary training at Pulaski High School, graduating in 1961. His undergraduate training was obtained at The University of Wisconsin at Milwaukee from which he received a Bachelor of Science degree in Psychology in 1965. On June 19, 1965 he married Claudine Semrad and they moved to Seattle, Washington. He started his graduate training at the University of Washington in the autumn of 1965.