EFFECTS OF PRIOR PROCESSING ON JUDGING GYMNASTICS

EFFECTS OF PRIOR PROCESSING ON JUDGING GYMNASTICS

By

DIANE STE-MARIE, B. Ed.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Master of Science

McMaster University
August, 1989

McMASTER UNIVERSITY LIBRARY

McMASTER UNIVERSITY LIBRARY

Master of Science (1989)
(Adapted Human Biodynamics)

McMaster University
Hamilton, Ontario

TITLE: Effects of Prior Processing on Judging Gymnastics

AUTHOR: Diane Ste-Marie, B. Ed. (McGill)

SUPERVISOR: Dr. Timothy D. Lee

NUMBER OF PAGES: vii, 76

Abstract

Three experiments are reported that examined the influence of prior judgements on implicit and explicit tests of memory in gymnastic judging. The rationale was that if gymnastic judging is affected by the memory for prior episodes, then the accuracy of judgements should change as a result of the prior episode. The results of Experiment 1 revealed that perceptual judgements differed as a function of an item's relationship between the study phase and the perceptual test phase. Moves that had the same performance in both phases resulted in the highest level of accuracy ($\underline{M}$ = 79%). New moves were less accurate ($\underline{M}$ = 75%). The lowest level of accuracy was achieved for items where the performance was altered between study and the perceptual test ($\underline{M}$ = 72%). Similarly, recognition judgements differed as a function of an item's relationship between the study phase and recognition test phase. Novice and expert judges revealed similar memory influences for perceptual and recognition judgements (Experiment 1). Memory influences were reduced, but still evident when subjects were given prior knowledge of these effects and procedural changes were adopted (Experiment 2). Spacing of repetitions did not enhance prior processing effects of perceptual judgements, but superior retention was noted for spaced repetitions in the recognition test phase (Experiment 3). These findings are discussed in terms of memory influences on subjective experience and the practical implications of judges' exposure to an athlete's performance prior to competition.

ACKNOWLEDGEMENTS

Thank you to my committee members; Larry Brawley, Larry Jacoby and Jan Starkes for their interest and support.

A special thanks is extended to Tim Lee for his continued insight and guidance. I have been fortunate in finding not only a great supervisor but a good friend.

This thesis is dedicated to my cherished husband, Steve, for his devotedness to learning and pursuing knowledge, and to my parents for their loving support and encouragement in continuing my education.

Table of Contents

List of Tables

Table		Page
21	Study phase cell means for Experiment 2	72
22	Perceptual test phase cell means for Experiment 2	72
23	Recognition test phase cell means for Experiment 2	73
24	Voice reaction time cell means for Experiment 2	73
25	Study phase cell means for Experiment 3	74
26	Perceptual test phase cell means for Experiment 3	74
27	Recognition test phase cell means for Experiment 3	75
28	Perceptual test phase voice reaction time cell means for Experiment 3	75
29	Recognition test phase voice reaction time cell means for Experiment 3	76

List of Figures

Introduction

A single prior exposure to a word has long lasting effects, both aware and unaware, on memory-influenced judgements. For example, subjects in Kolers' experiment (Kolers, 1976) showed a savings for speed of reading typographically transformed text from passages that had been read over a year earlier although they could not explicitly remember the passages. Similar memory effects on perception and recognition judgements have led to the suggestion that perception is heavily influenced by a single prior processing episode (e.g., Eich, 1984; Jacoby, 1983a, 1983b; Witherspoon & Allan, 1985). Moreover, these effects are rather robust. Memory influences have been shown for perception of non-words, sentences, pictures, and for stem- and fragment-completion tasks (see Richardson-Klavehn & Bjork, 1988, for a review). Collectively, these effects cast serious doubt on the contention that perceptual judgements are based on abstract representations that are not biased by prior processing episodes.

The generality of these effects raises concern for sports such as gymnastics where perceptual judgements serve a dominant role. The judge's task in gymnastics is to watch a very brief display and make a judgment based upon the perception of that display -- a task not dissimilar to the subject's task in a perceptual identification experiment (see Salmela, 1978 for further discussion of the task demands). If the effects of prior processing episodes on perceptual judgment shown for laboratory tasks also occur in gymnastics, then the nature of the performance in the prior exposures might be expected to bias the perception of the competitive performance. Further, this bias could be either beneficial or detrimental to the performer's score. For instance, if the performer makes an error during the warm-up but not during the competition, the judge's memory for that error might bias the perception of the better performance during competition. That bias would likely result in a score that was lower than an unbiased assessment. Conversely, if the warm-up was performed well but the competitive performance was flawed, the bias might result in awarding a higher score than the performance deserved.

Gymnastic judges served as subjects in the present series of experiments to investigate this issue. These experiments were designed to parallel the procedures used in previous laboratory research (Jacoby & Dallas, 1981, Expt 2a), and also to simulate the warm-up/competition setting encountered at gymnastic competitions. Thus, the design involved three phases for each experiment. In the study phase, judges watched a series of edited gymnastic moves and decided whether or not the move contained a form error. The judge's task was the same for the perceptual test phase, except that some of the moves had been seen earlier during the study phase -- either with the same performance or with a different performance. Following the perceptual test phase the judges viewed a further series of moves and made recognition decisions about whether that performance had been seen during the study phase.

Judges use different strategies to overcome the complexity of the information processing demands encountered at a gymnastics competition (Salmela, 1978). For example, some judges use the within team order as a factor to aid them in determining the athlete's score (Ansorge, Scheer, Laub & Howard, 1978). Gymnasts who competed last in a rotation received higher scores than if they had competed first in the rotation. Another strategy to assist in judging is the information used from previous warm-up routines (Salmela, 1978). To our knowledge however, there is no literature available that investigated this suggestion, even though there is considerable evidence from laboratory tasks to support this suggestion.

Memory influences have been shown to bias subjective judgements (Jacoby, Allan, Collins & Larwill, 1988; Witherspoon & Allan, 1985). For example, words that had been previously presented were judged as having stayed on the screen longer than words that had not been previously presented (Witherspoon & Allan, 1985). That prior episodes affect subjective judgements sheds light on the potential effects that prior exposures might have on gymnastic judging. While watching a previous episode might reduce the information processing demands of the judge, memory for the prior episodes might also bias the subjective judgements made by the judges.

General Methodology

Since the three experiments reported here are similar, the methods that were common to all studies will be described. Variations in the general method, and details of the specific design and stimuli for individual experiments will be indicated as each experiment is presented.

Subjects. Female gymnastic judges, certified by the Canadian Gymnastic Federation, volunteered to participate in these experiments. Subjects were recruited at various gymnastic competitions in the southern and central Ontario regions. Each judge participated in only one experiment.

Materials. A Sony video eight camera (model CCD-V9/V90) was used to film the gymnastic moves. These were performed by four gymnasts; two were national-level competitors and two were university-level competitors (all were dressed similarly, in dark bodysuits, during the filming). Although there were four gymnasts, both versions of any one move were performed by the same gymnast. Each gymnastic move was performed many times for later review. As a result of the review, two versions of each move were selected to be used as stimuli in the experiments: 1) performance perfect, and 2) performance with a form error[1] . When the move was performed with a form error only one form error occurred.

The selected video segments were then edited in 1 s video clips onto a Sony beta video cassette[2]. This resulted in a portion, rather than the entire gymnastic move being presented[3]. The editing of the moves was done such that the same tape frames were shown both for the perfect performance and the form error performance[4]. There was an equal number of moves for each of the four events in which women compete (vault horse, uneven bars, balance beam, and floor exercise), and these were ordered randomly throughout the phases. For each of the phases half of the moves were demonstrated perfectly and half were demonstrated with a form error. Video clips were projected from a Sanyo Betaccord (model # 4650) on a 16 inch RCA XL-100 television screen. Subjects were seated approximately 1 m from the television but were free to adjust this distance.

Procedures. The experiment was conducted in three phases: 1) a study phase 2) a perceptual test phase and 3) a recognition phase.

Study Phase. This phase consisted of a number of formats dependent on the number of conditions in the experiment. Subjects were assigned to one of the formats, in a counterbalanced order. Subjects watched a series of gymnastic moves interspersed with blank screen intervals. After the presentation of each gymnastic move the subject reported whether the move had been performed perfectly or with a form error. If a form error had been detected, the judge was to specify the nature of that error.

Perceptual Test Phase. All subjects were administered the same format in this phase, thereby resulting in all subjects viewing the same perceptual test stimuli. Thus, if there were any changes in judgements from the study phase to the perceptual test phase, the move itself could not be considered the determining factor. Rather, any differences in judgements would be due to the nature of the prior exposure.

The subjects watched a series of gymnastic moves and reported their judgements after each move was presented. The gymnastic moves were of three critical types, defined in terms of an item's relationship between the perceptual test phase and the study phase; "same", "different" and "new". Moves that had no change in performance from study to test were considered "same" (in fact, "same" items were the actual video clips that had been presented in the study phase). Moves that had changed in performance from study to test were considered "different". In these instances, a video clip with the same taped segment of a move presented in study phase was presented in the perceptual test phase, but with the alternate performance outcome. For example, if the performance was perfect in the study phase, it was performed with a form error in the perceptual test phase (and conversely, a form error in the study phase was performed perfectly in the perceptual test phase). "New" moves were those that had not been presented in the study phase[5].

Recognition Test Phase. Similarly, this phase also had only one format presentation (for the same reasons described in the perceptual test phase section). In this phase, subjects watched gymnastic moves and were asked to make an old/new decision after each. If the performance of a move in the recognition phase was the same as that in the study phase ("same"), the correct response was old. However, if the performance in the recognition phase differed from the performance in the study phase ("different"), the subject was to judge the move as new. If the move had not been seen in the study phase ("new"), the move was also to be judged as new. Subjects were thanked and fully debriefed at the conclusion of the experiment.

The subjects responded verbally for all three phases. The experimenter stayed in the room with the subjects to read a set of instructions at the beginning of each phase and to record the subjects' responses. Since the subjects were not recruited from the same area, the experimental sessions were not in the same room for all of the subjects. However, all subjects were tested individually by the same experimenter. An experimental session lasted approximately 30 min.

The dependent measure for the study and perceptual test phases was the accuracy of each response (reported as percent correct). The recognition scores were analyzed as the probability of an item being judged as "old". For Experiments 2 and 3, time to respond (VRT, voice reaction time) was also included in the perceptual and recognition test phases. Unless otherwise indicated, the significance level was set at .05 for all the statistical tests. Neuman-Keuls post hoc tests were used to determine significant differences between means. Statistical models for each analysis of variance (ANOVA) will be described later.

Experiment 1

Word frequency is one of the most important variables to affect perceptual recognition (Jacoby & Dallas, 1981). High frequency words are recognized more easily than words that occur less frequently (Murrel & Morton, 1974). However, one prior presentation is sufficient to reduce the advantage that high frequency words have over low frequency words (Jacoby & Dallas, 1981; Scarborough, Cortese, & Scarborough, 1977). Thus, low frequency words benefit more from a prior presentation than do high frequency words. Of interest to the present experiment, was to determine if a similar effect would occur as a function of expertise in gymnastic judging. To investigate this, an additional factor distinct to this experiment, was the testing of both expert and novice judges as subjects. Novice judges were expected to be influenced more by a prior presentation of a gymnastic move than were expert judges. The analogy here is that in experiments where words are used as stimuli, subjects are novices with some words -- words they have not had much exposure to (i.e low frequency words), and expert with other words -- those they have had much exposure to (i.e. high frequency words). Similarly, novice judges (as compared to expert judges) have had less exposure to the variety of gymnastic moves, and more specifically to the potential errors that could occur in any given move. Hence, the hypothesis was that experience would improve the objectiveness of a gymnastic judgement, and that objectiveness will be less biased by a single prior presentation for experts than for novices

Much research has been devoted to the topic of expertise. As a result, many findings have emerged on differences between novice and experts across various domains (see Glaser & Chi, 1988 for a review). Of particular interest to the present experiment is the evidence that performance on analytic-like sport tasks improve as a function of increased experience (Allard & Starkes, 1980; Biscan & Hoffman, 1976). For example, error detection rate for the forehand serve in tennis was superior for experienced tennis coaches than inexperienced subjects (Armstrong & Hoffman, 1979). Thus, it was expected that expert judges would demonstrate superior error detection capabilities than novice judges.

Method

Subjects. Twenty-four subjects were divided into two groups based on their years of experience judging gymnastics. Subjects with one to three years of experience were classified as novice judges (*M* years of experience = 1.6, range = 1 - 3; *M* age = 26.8, range = 17 - 45). Subjects with greater than ten years experience were classified as expert judges (*M* years of experience = 13.2, range = 10 - 19; *M* age = 35.2, range = 28 - 43).

Stimuli. Stimuli consisted of 256 gymnastic moves. Of these, 144 were critical items and the other 112 were used as lead-ins and fillers throughout the three phases. The 144 critical items were 72 gymnastic moves performed twice each; once perfectly and once with a form error. These were divided into three equal sections: 24 performed perfectly, 24 performed with a form error, and 24 that were not shown in the study phase of the experiment. These gymnastic moves were counterbalanced across subjects so that each move served in each of the three blocks equally often. Thus, three formats were constructed for the study phase. For example, if, in format #1 a subject viewed a cartwheel performed "perfectly", a subject in format #2 viewed the cartwheel with a "form error" and a subject in format #3 did not view the cartwheel at all. Subjects were assigned to one of the formats such that four novice and four experts served in each format.

Procedures

Study Phase. Subjects watched a series of 80 gymnastic moves interspersed with 5 s blank screen intervals. Of these, 8 were lead-ins (to familiarize the subject with the task), 24 were fillers (to allow for balancing of items that were not to be seen), 24 were critical items performed perfectly, and 24 were critical items performed with a form error. The 72 critical items were distributed evenly throughout each format. Emphasis was stressed on the accuracy of each decision. There was no emphasis on speed of response. Preliminary testing indicated that the 5 s interstimulus interval was sufficient for the judges to respond and prepare for the next presentation.

Perceptual Test Phase. All subjects were administered the same format. Subjects watched 64 gymnastic moves and reported their judgements following each item. Of these, 4 were lead ins, 16 were fillers, and 48 were critical items. The 48 critical items were comprised of the three types; 16 "same", 16 "different", and 16 "new". The time between study and the perceptual test phase was approximately 2 min.

Recognition Test Phase. This final phase was an unexpected recognition test. Subjects watched 40 gymnastic moves and made an old/new decision after the presentation of each move. Of these, 8 were lead-ins, 8 were fillers, 8 were "same", 8 were "different", and 8 were "new". All of the "same" moves were from the study phase (none of these moves were shown in the perceptual test phase). The "different" moves were related to the study phase, in that all the moves had a counterpart in the study phase that had a different performance (none of the moves were shown in the perceptual test phase). The blank screen interval between each gymnastic move in the recognition phase was approximately 2 s.

Results

Study Phase. The accuracy scores were analyzed using a 2 (group: expert/novice) x 2 (performance: perfect/error) ANOVA with repeated measures on the last factor. Perfect performances were detected equally well by novices (M = 85.4%) and experts (M = 84.9%). However, experts (M = 77.7%) detected error performances better than novices (M = 64.5%). This interaction of skill level and performance was significant, $F(1,22) = 5.91$, $MSe = 93.45$. In addition, main effects for skill level, $F(1,22) = 7.54$, $MSe = 64.16$ and performance, $F(1,22) = 25.30$, $MSe = 93.45$ were also found.

Perceptual Test Phase. The accuracy scores were analyzed using a 2 (group: expert/novice) x 2 (performance: perfect/error) x 3 (type: same/different/new) ANOVA with repeated measures on the last two factors. The accuracy of judgements differed as a function of the move's relationship between the study phase and the perceptual test phase. Gymnastic moves that had the same performance in both phases were judged most accurately (M = 78.7%). New moves were judged less accurately (*M* = 75.4%). The lowest level of accuracy occurred for items where the performance outcome was different between study and test (*M* = 72.1%). These differences were supported by a significant main effect for type, *F* (2,44) = 3.91, *MSe* = 14.74. A Neuman-Keuls test indicated that the "same" moves differed significantly from the "different" moves. The "new" moves were not significantly different from either of the other types.

Insert Table 1 about here

Again, experts were more accurate than novices in detecting errors (73.4% vs. 59.4%) but not in detecting perfect performances (81.3% vs. 87.4%). This interaction of skill level and performance was significant, *F*(1,22) = 6.43, *MSe* = 569.62, as was the performance type main effect, *F*(1,22) = 20.32, *MSe* = 569.62.

Since no interaction was found between skill level and type of item (*F* < 1.0), the data were collapsed across skill level and are presented in Table 1. The absence of an interaction reveals that the influence of a prior exposure on judging form errors was similar for both experts and novices.

Recognition Test Phase. The probability of judging an item as old was analyzed using a 2 (group: expert/novice) x 2 (performance: perfect/error) x 3 (type: same/different/new)

ANOVA with repeated measures on the last two factors. Since no interaction was found between skill level and type of item ($F < 1.0$) the data for the recognition phase were collapsed across skill level and are presented in Table 2.

The probability of judging an item as old differed as a function of the relationship between the recognition item and an item in the study phase. The highest level of probability occurred for moves that were exactly the same in both phases (M = 69.3%). New moves were also judged accurately, having a low probability of being judged as old (M = 28.3%). However, items that differed in the performance outcome from the study phase to the recognition phase were poorly identified (M = 48.6%). These differences were supported by a main effect for item type, $F(2,44)$ = 26.39, MSe =764.53. The post-hoc tests revealed that all pairwise differences were significant. Also noted here was the absence of either a group main effect or an interaction of group and type of move (F's < 1.0). As with the perceptual test phase, experts and novices showed similar influences of prior processing on memory performance.

Insert Table 2 about here

Discussion

Although experts were better than novices in correctly detecting the occurrence of form errors, the influence of a prior episode on both perceptual judgment and recognition memory was similar. Both novice and expert judges perceived moves that were seen for the second time, with an identical performance, with greater accuracy (M = 78.7%) than when the move was seen for the first time (M = 75.4%. As well, moves that changed performance on the second viewing, were perceived with less accuracy (M = 72.1%) than when seen for the first time. This evidence weighs against an interpretation of skill at judging gymnastics as the development of a memory system which becomes increasingly abstract.

Clearly, the expert judges performed better than novices at doing the task that suits their expertise (identifying form errors). Also, that no group differences were found on the recognition task suggests that the judge's skill is relatively specific. However, to find similar influences of prior processing episodes in both groups of judges implies that the prior processing of a single specific event has lasting effects in memory, and that this memory cannot be easily discounted when later performing the same task. Perceptual judgements of aesthetics in gymnastics appears to be subject to memory influences regardless of the skill level of the judge.

The perceptual test phase of the experiment revealed that detecting the occurrence of form errors was affected by an earlier prior processing. When the performance of a move changed from that of the first viewing, the accuracy of the judge's assessment was relatively poor. Moves that had not been observed before were scored more accurately and repeated performances were assessed most accurately.

The recognition scores indicated that the judges were good at recognizing moves that had not been seen during study (M = 28.3%) as well as moves that had been seen exactly as during study (M = 69.3%). However, recognition of moves for which the performance outcome had changed was poor (49%). Since the judges were poor at detecting changes in performance from study to recognition, the implication is that the effects seen during the perceptual test phase were unintentional biases due to memory for a prior processing episode.

There are two factors that potentially confound the above interpretation: 1) stimuli were not the same for perceptual and recognition test phases and 2) the interval of time following the study phase were different for the perceptual and recognition tests. The first confound arises as a result of not being able to generalize between phases. Items in the perceptual phase were not the same as those in the recognition phase. Thus, the assumption that poor recognition in the recognition test phase presupposes poor recognition of the perceptual test phase items is only warranted if items specific to each test , did not play a role.

The second confound was that the recognition phase always occurred after the perceptual test phase. The retention interval for the recognition test was approximately 25 min, whereas the retention interval for the perceptual test was approximately 2 min. This constant ordering of test phases could have produced carry over effects: the perceptual test phase material may have interfered with the study phase material, resulting in increased difficulty for the recognition judgements. Interference effects have been typically used to account for forgetting in recognition tests (Bowles & Glanzer, 1983; Underwood, 1957), and may apply to this experiment as well.

The rationale for the constant ordering of the phases was that if the recognition phase had been placed before the perceptual test phase, the subject may have approached the perceptual test differently than if it had not been preceded by the recognition phase (Jacoby, 1988). In all likelihood, the subject would have treated the perceptual test more analytically, first deciding if the move was new or old, and if old, what decision had been made previously during study. Since the purpose of the perceptual test was an implicit assessment of memory influences on perception, the constant ordering procedure best suited our needs.

Experiment 2

The findings of Experiment 1 have both practical and theoretical significance. A further issue of both practical and theoretical importance is whether or not prior processing influences can be eliminated by subjects' awareness of these effects. Jacoby and Kelley (1987) stated that upon awareness and understanding of a prior event influencing our current perceptions and judgements, we can deliberately change the basis of our judgements and escape the influences of that prior event. Several changes in the basic methodology were made to investigate this possibility.

The key difference of the present experiment was that subjects were made aware of the findings of Experiment 1 prior to the study phase. In the instructions here, subjects were told that the findings of a previous experiment suggested that judges may be influenced by their previous judgements. Subjects' prior knowledge of the influences may cause a shift to a more analytical basis for the performance judgements (Jacoby & Kelly, 1987). That is, the judge would consider if the move was seen before, and if so, with what performance. Thus, subjects were told that each move was to be assessed with the knowledge that there is a comparison to be made. This comparison may serve to eliminate the influence of the prior viewing of the gymnastic move. As well, judges were informed that there were two phases in the experiment, and then cautioned to try not to allow the decisions they made in the study phase to influence their judgements in the second phase. These changes served to make the subjects aware of prior processing influences and also emphasized avoidance of the effects. The hypothesis was that knowledge of the effect would reduce, if not eliminate, the influence of prior processing on later judgements.

Another change in the present experiment was that the perceptual and recognition test phases were combined into a single test phase. This change served to address the two potential confounds from Experiment 1. First, both perceptual and recognition judgements were made on the same items. Second, the retention intervals for the perceptual and recognition judgements were no longer of varying lengths. Thus, an interpretation of the perceptual data based on the performance of the recognition data for the same items was possible.

Increased speed of identification has been another measure used to reveal memory influences (Jacoby & Dallas, 1981; Scarborough, Gerard & Cortese, 1979). Responses to repeated items are quicker than responses to items seen for the first time. To examine if this memory influence also occurs when judging gymnastics, an additional dependent measure of voice reaction time (VRT) was included in Experiments 2 and 3. Same items were expected to be judged quicker than items that had changed in performance.

Method

Subjects. Fifteen female gymnastic judges volunteered to participate in this experiment. Since novice and expert judges had similar influences for both memory tests in Experiment 1, no restrictions were made for experience.

Materials. A tone was recorded on tape prior to the presentation of each gymnastic move. The tone initiated a millisecond timer (Lafayette 54417-A) which was stopped by a vocal response made into a microphone (Realistic model MC-1000). The experimenter recorded both the subject's response and the VRT. The VRT scores were collected only during the test phase.

Stimuli. Stimuli consisted of 176 edited gymnastic moves. Of these, 56 were lead-ins and fillers, the remaining 120 were critical items. These 120 items included 60 gymnastic moves edited with a perfect performance and the same 60 moves edited with a form error. The 60 gymnastic moves were counterbalanced across three blocks of 20, following the same procedure as Experiment 1. Three formats were created for the study phase and subjects were assigned at random to one of these (with the restriction of 5 subjects per format).

Procedures

Study Phase. Subjects watched a series of 68 gymnastic moves, interspersed with 4 s blank screen intervals, and reported their judgements following each move. Of these, 8 were lead-ins, 20 were perfect performances, 20 were error performances, and 20 were fillers. Instructions stressed the importance of both speed and accuracy. Although VRT was not recorded here, subjects spoke into a microphone and were told that speed and accuracy were being recorded.

Test Phase. The instructions in this phase differed from that in Experiment 1 in two respects. The introduction of the VRT scores as a dependent measure required that subjects were responding with the same intentions (i.e. all responding to form error first). To this end, subjects were instructed to first report their form error judgements as quickly and as accurately as possible. Immediately following the form error judgement the subject was to make an old/new decision. Moves were to be called old only if the video clip was identical to one from the study phase ("same"). Moves with a different performance ("different") and moves that had not been seen in the study phase ("new") were to be called new. Speed of response was not stressed in the old/new decision (only accuracy).

The second difference was that the instructions included two pieces of information that subjects in the first experiment did not receive. In this experiment, subjects were informed of the results of a previous experiment and the possible practical implications of these results. Further, subjects were told they were going to judge gymnastic moves in two separate phases and to try not to allow the judgements made in the first phase to influence their judgements in the second phase.

Eighty-eight gymnastic moves were presented, interspersed with 6 s blank screen intervals. Of these 8 were lead-ins, 20 were fillers and the remaining 60 were critical items. The critical items were equally divided into "same", "different", and "new" item types.

Results

Study Phase. The accuracy scores (reported as percent correct) were analyzed using a one way repeated measures ANOVA (performance: perfect/error). The difference between perfect (M = 83%) and error (M = 76.7%) performances was not significant, $F < 1.0$.

Perceptual Test. The scores (reported as percent correct) were analyzed using a 2 (performance: perfect/error) x 3 (type same/different/new) ANOVA with repeated measures on both factors. Gymnastic moves that had the same performance in both phases resulted in the highest level of accuracy (M = 71.2%). New moves were judged less accurately (M = 68.2%). The lowest level of accuracy was achieved for items where the performance had differed between study and test (M = 66.2%). Although the differences among these means did not reach significance, $F(2,28) = 1.25$, $MSe = 149.27$, the direction of the results followed the trend that was found in the initial experiment. No other effects reached significance. The voice reaction time data failed to add any new information, (all F values less than 1.0).

Recognition Test. Recognition data were analyzed in terms of the probability of judging an item as "old" (having been presented in the study phase). The scores were analyzed using a 2 (performance: perfect/error) x 3 (type: same/different/new) with repeated measures on both factors.

The probability of judging an item as old differed as a function of the relationship between the recognition item and an item in the study phase. The highest level of probability was achieved for items that were exactly the same in both phases (M = 64.1%). New moves were also accurately judged, having a low probability of being called old (M = 26.2%). However, moves that differed in performance from study to test were not recognized as well (M = 48.5%). These differences were substantiated by a main effect for item type, $F(2,28) = 34.4$, $MSe = 315.27$. Neuman Keuls post-hoc tests indicated that each of the types of items differed from the others. No other effects reached significance. The VRT data failed to add any new information, (all F values less than 1.0).

Discussion

Effects of prior processing were reduced, but perceptual biases were still evident. This reduction of the effect was presumably related to the procedural changes adopted in the present study. Recall that the subjects were cognizant of the effects noted in the previous experiment. This knowledge was complemented with the judges being alerted to not let their judgements be affected by a prior viewing. Thus, awareness of the potential influences decreased the effect of prior episodes on the judging of gymnastics, as was hypothesized. However, it is difficult to say (both statistically and practically) that a prior viewing of an episode had <u>no influences</u> based on these data. The ordering of the means was the same as in Experiment 1, the only distinction being that the magnitude was reduced slightly in this experiment. Rather than stating that there was no effect of having seen a prior performance of a gymnastic move, a more prudent interpretation of the data was that the effect was diminished.

The supplementary information preceding the experiment was not the only difference between the first and second experiment. Also included was the combination of the perceptual and recognition judgements into one phase. Reduction of the effect may also be attributed to the differences in the processes involved when subjects perform both tasks within the same phase (Jacoby, 1988; Witherspoon & Moscovitch, 1989). Impromptu comments by subjects indicated that they felt it was difficult to ignore the old/new decision before responding to the performance judgements. However, an implicit test is to be performed without attention being given to the past (Graf & Schacter, 1985). It is apparent that this was difficult to attain given that subjects most likely made both decisions concurrently. This change in strategy contaminated the assumption of an implicit test, and may also have reduced the effects previously produced by that implicit test.

In the first experiment, there were two problems that confounded an interpretation of the perceptual test data based on the performance in the recognition phase. The first confound was that the items in the perceptual test phase were not the same as the recognition test phase items. The second concern was that the experimental design resulted in an increased retention interval for the recognition test phase. As well, items from the perceptual test phase may have interfered with items from the study phase. Despite the change in the present experiment of combining the perceptual and recognition test phases into one test phase, the previous recognition test findings were replicated. Judges were able to recognize moves that had been seen exactly as before (M = 64.1%) and identify those that had not been shown in the study phase (M = 26.2%). However, being able to determine if the move differed in performance proved to be a difficult task. Recognition of these moves was poor (48.5%), indicating that judges were poor at detecting changes in performance from study to test. This replication of the recognition data strengthens the contention that the effects seen on perception are unintentional biases due to memory for a prior processing episode.

The combination of the two phases also enabled an examination of "different" perceptual test items as a function of whether or not it had been explicitly remembered in the recognition phase. If the prior processing influence was <u>not</u> due to unintentional biases, then it was expected that incorrect perceptual judgements for the "different" items would have occurred only for the incorrectly identified items (those called old). The assumption here is that if the "different" item is correctly called new, then the judge must have realized the performance had changed, and thus would have changed her form error judgement accordingly (resulting in no bias). However, if the judge does not change her form error accordingly, then the judge's bias in this instance would be unintentional. That is, the judge was aware of the change in performance , but unable to discount the influences of the prior event.

To conduct this analysis, perceptual judgement accuracy was analyzed based on the items' recognition judgements (items called "old" as compared to items called "new"). One-way ANOVAs were used to determine if the effects on perception were a function of explicitly remembering an item. If recognition mediated perception, then the "different" items correctly identified as new were expected to be perceptually judged more accurately than those incorrectly identified as old. However, the decision of old versus new had no differential effect on the accuracy of perception. Items correctly called new ($\underline{M}$ = 66.9%) experienced the same amount of influence as items called old ($\underline{M}$ = 65.5%). That is, realization of a change in performance nevertheless resulted in biased judgements. Thus, the premise that the judges' awareness of a change in performance presupposes a correct perceptual judgement was not supported. Instead, the situation arises where even though the judge acknowledges that the performance has differed, the judge was influenced by the prior episode. Hence, further support is provided for the conclusion that unintentional influences undermine accurate perceptual judgements in gymnastic judging. Thus, it appears that judges were unable to discount the influence of the prior exposure on their perceptions.

This finding of unintentional biases is comparable to the hindsight effect (Fischoff, 1975). Fischoff (1975) reported that once subjects were given the outcome of a certain event, subjects overestimated what they would have known without the outcome knowledge. More surprisingly, subjects could not eliminate this hindsight effect, even when given instructions explaining their biases. Similarly, gymnastic judges were influenced by an outcome of a previous event. The influences of the prior event continued to persist despite judges' awareness of the potential judgemental biases.

The present experiment provided additional evidence that prior episodes affected judges' assessments. And, that unintentional influences are a source of the prior processing effects. Parallel effects on perception and recognition memory by prior episodes were also maintained in this experiment.

Experiment 3

The ordering of highest to lowest accuracy levels for "same", "new", and "different" items respectively, revealed judges' biases for memory of a prior event in Experiment 1. In the second experiment, prior knowledge of the effect and the combination of the two phases reduced the effect, but the pattern of accuracy levels was consistent with that of Experiment 1. The purpose of the present experiment was to determine if the prior processing effect could be enhanced.

Jacoby and Dallas (1981, Expts. 4a,4b) reported that two presentations of an item enhanced perceptual and recognition memory, as did the increased spacing of the repetitions. Different measures of implicit and explicit tests have since replicated these enhanced and parallel effects between the two types of tests (Feustal, Shiffrin & Salasoo, 1983; Graf & Mandler, 1984). The study-test procedure implemented by Jacoby and Dallas (1981, Expts 4a,4b) was adopted in this experiment. Gymnastic moves in the study phase were presented once, or twice massed (adjacent presentations), or twice spaced (ten moves intervening presentations). It was hypothesized that observing a move performed twice with the same performance, prior to a third viewing with the same performance, would enhance the accuracy of perceptual and recognition judgements (relative to just one prior processing). However, if the third viewing was different from that of the first two then perception and recognition memory would be poorer than just following one prior presentation. Further, these repetition effects on perception and recognition memory were expected to be greater for spaced than for massed items during study (Jacoby & Dallas, 1981, Expts 4a,4b; Scarborough, Cortese, & Scarborough, 1977; Scarborough, Gerard, & Cortese, 1979).

Notably, in the second experiment subjects mentioned having difficulty reporting the form error judgements without thinking about what they had seen before. However, an implicit task is to be performed without reference to a prior episode (Graf & Schacter, 1985). To adhere to this criteria it was decided to return to the procedures of Experiment 1. Hence, the recognition phase was administered following the perceptual phase.

Method

Subjects. Fourteen judges volunteered to participate in this experiment. No restrictions were made with reference to judging experience.

Stimuli. Stimuli consisted of 208 gymnastic moves. Of these, 168 were critical items and the other 40 were used as lead-ins and fillers throughout the three phases. The 168 critical items were 84 gymnastic moves performed twice each; once perfectly and once with a form error. These were divided into seven equal sections: 12 perfect performances seen once, 12 error performances seen once, 12 perfect performances seen twice (massed), 12 error performances seen twice (massed), 12 perfect performances seen twice (spaced), 12 error performances seen twice (spaced), and 12 items which were not seen in the study phase. These gymnastic moves were counterbalanced across subjects so that each move served in each of the seven sections equally often. The procedure necessitated that seven formats be constructed for the study phase. Subjects were assigned to one of the seven study formats at random but with the restriction that there were two subjects in each format.

Procedures

Study Phase. Subjects watched a series of 136 gymnastic moves interspersed with 4 s blank screen intervals. Of these, 4 were lead-ins, 12 were fillers, and 120 were critical items. These critical items were; 12 error performances seen once, 12 perfect performances seen once, 12 error performances seen twice massed, 12 perfect performances seen twice massed, 12 error performances seen twice spaced, 12 perfect performances seen twice spaced.

The items that were seen twice in this phase had the same performance on both occasions. Upon presentation of each video clip the subject reported if the move had been performed perfectly or with a form error.

Perceptual Test Phase. All subjects watched the same format (which consisted of 68 gymnastic moves) and reported their judgements after each video clip. Of these, 4 were lead-ins, 8 were fillers and 56 were critical items. Eight of the critical items were the item type "new", 24 were the item type "same", the remaining 24 were "different". The interval between the study phase and this phase was approximately 2 min.

Recognition Test Phase. The final phase was an unexpected recognition test. In this phase subjects viewed 40 gymnastic moves and were asked to make an old/new decision following each presentation. Moves that were identical to one in the study phase ("same") were to be called "old". Moves that were the item type "new" or "different" were to be called "new". All of the moves in this phase were from the study phase (none of the moves had been presented in the perceptual test phase). Of the 40 items, 4 were lead-ins, 8 were fillers, 4 were "new", 12 were "same", and 12 were "different". The retention interval from study to the recognition test was approximately 25 min.

Results

Study Phase. The study phase results are summarized in Table 3. Study phase scores were analyzed using a 2 (performance: perfect/error) x 5 (type: once seen/first massed item/second massed item/first spaced item/second spaced item) ANOVA with repeated measures on both factors. Since the design of the experimental factors was an incomplete factorial, the ANOVAs were followed by planned comparisons.

The planned comparisons revealed two interesting findings. Firstly, items that were seen twice (M = 78.1%) were more accurately judged than items seen once, (M = 73.6%), $F(1,13) = 5.52$, $MSe = 121.93$. Also, with error performances, the second presentation of an item (M = 77.3%) was more accurately judged than the first presentation (M = 65.6%) for spaced items, $F(1,13) = 9.72$, $MSe = 99.13$. However, with massed items the accuracy level of error performances did not differ between the second presentation, (M = 73.3%) and the first presentation, (M = 72.8%). Again, the difference between judging perfect performances (M = 80.1%) and error performances (M = 70.7%) was not significant, $F(1,13) = 2.83$, $MSe = 1311.89$.

Insert Table 3 about here

Perceptual Test Phase. Since the experimental design was an incomplete factorial, moves that were seen for the first time ("new") were omitted from the initial set of perceptual test phase analyses. Instead, the means from this item type will be used as a control value. Data were analyzed using a 2 (performance: perfect/error) x 3 (repetitions: once/twice massed/twice spaced) x 2 (type: same/different) ANOVA with repeated measures on all factors. The data for the perceptual test data are presented in Figure 1.

Items seen once with the same performance at study and test ("same", M = 78.9%) were more accurately judged than items that changed performance ("different", M = 67%). A similar effect occurred for spaced items: "same" item type (M = 83.7%) were judged more accurately than and "different" items (M = 72.4%). However, "same" massed items (M = 76.9%) were not different from "different" massed items (M = 79.5%). This interaction of repetition and type was statistically significant at p = .055, $F(2,26)$ = 3.21, MSe = 292.60. In addition, the main effect for type was also statistically significant p = .055, $F(1,13)$ = 4.31, MSe = 461.35. The VRT data added no new information, (all F values less than 2.0).

Insert Figure 1 about here

Recognition Test Phase. Similarly, the recognition test phase also excluded the "new" item type from the recognition test phase analysis. The mean for this item type will be included as a control value. Recognition scores were analyzed using a 2 (performance: perfect/error) x 3 (repetition: once/twice massed/twice spaced) x 2 (type: same/different) ANOVA with repeated measures on all factors. The recognition test data are presented in Figure 2.

Perfect and error performances of the "same" items that were spaced (perfect, *M* = 75%; error, *M* = 82.1%) were recognized better than perfect and error performances of the "different", spaced items (perfect, *M* = 50%; error, *M* = 39.3%). Perfect performances of the "same" massed items (*M* = 75%) were recognized better than the perfect performances of the "different" massed items (*M* = 32.1%). However, no differences occurred between error performances of the "same" massed items (*M* = 64.3%) and the "different" massed items (*M* = 50%). Recognition of the perfect and error performances of the items seen once did not differ as a function of item type. This three way interaction of performance, repetition and type was significant, $F(2,26) = 4.49$, $MSe = 559.75$. Main effects for repetition, $F(2,26) = 3.30$, $MSe = 1111.49$, and item type, $F(1,13) = 20.71$, $MSe = 1149.27$ were also found.

Insert Figure 2 about here

The VRT scores indicated that items that were "different" ($\underline{M}$ = 2.87 s) took longer to respond to than the "same" items ($\underline{M}$ = 2.73 s). This difference was supported by a main effect of item type, $\underline{F}(1,13) = 5.40$, $\underline{MSe} = .145$. No other significant effects were found for VRT, (all $\underline{F}$ values less than 1.3).

Discussion

Many theories have emerged in the attempt to explain the spacing effect (see Hintzman, 1974 for a review). Melton (1967) related the spacing effect as a seemingly contradictory situation where forgetting during study appeared to improve memory. Cuddy and Jacoby (1982), interested in the paradox of Melton's suggestion, investigated the effects of forgetting on the magnitude of spacing effects. Cuddy and Jacoby elaborated on the "forgetting" hypothesis (see also Jacoby, 1978), stating that when the spacing of items resulted in forgetting, there was more complete processing of the second presentation, resulting in enhanced retention. However, when the second presentation of the item immediately followed the first, much of the processing was not repeated. Thus, massed items did not receive the same retention benefits as did spaced items.

Results of the study phase data in this experiment provide support for the forgetting hypothesis. A gymnastic move that was shown twice in succession was apt to receive the same judgement on both presentations, regardless of the accuracy of the judgement. For example, if the move had a form error that was not detected on its first presentation, the probability was low of detecting it on the second presentation (percent difference between first and second presentation of error performances for massed items, $\underline{M}$ = -0.5%). However, if the second presentation followed after ten intervening moves there was a higher probability of correct detection ($\underline{M}$ = 11.7% difference in improvement between first and second presentation of error performances for spaced items). It appeared that the second presentation of massed items (error performances) were incompletely processed (the judge not reexamining the performance). The second presentation (error performances) of a spaced item however, was likely examined with a more critical eye, and hence more accurately judged.

Although the perfect performances did not reflect this improvement in judgement accuracy from the first to the second presentation with spaced items, possible ceiling effects may account for this null effect. The accuracy level for first presented, spaced, perfect performance, items was already quite high ($\underline{M}$ = 81,5%). There doesn't appear to have been much room for improvement, especially, if the spaced items' perfect performances ($\underline{M}$ = 81.5%) were compared to the other means of the experiment, where only one value reaches a higher accuracy level ($\underline{M}$ = 85.1%).

Similar to the previous experiments, perceptual judgements were affected by the <u>single</u> prior processing of a gymnastic move. That is, the accuracy of the perceptual judgements for items that were seen once followed the same pattern as the preceding experiments. An increase in the magnitude of the difference between "same" and "different" items was expected to occur for items seen twice in the study phase, with the difference for spaced items being larger than the massed items. However, the results of the spaced and massed items did not support the proposed hypothesis. The difference between "same" and "different" items were of equal magnitude, 11.9% and 11.3% for once seen and spaced items respectively.

A number of studies have demonstrated that spacing between repetitions exerts parallel effects on implicit and explicit tests (see Richardson-Klavehn & Bjork, 1988 for a review). This experiment however, revealed a dissociation between the two tests. The explicit test followed the expected pattern of subjects' superior retention for spaced items over massed and once seen items. However, the implicit test did not follow the same pattern. Judges did not have superior perceptual test performance (or inferior, depending on the item type) for spaced items over massed and once seen items. Failure to demonstrate parallel effects of spacing of repetitions between perceptual and recognition memory for gymnastics judging was unexpected, although, not without precedent (Perruchet, 1989).

Perruchet (1989) completed four experiments examining the spacing effect in implicit (perceptual clarification procedure) and explicit (recall and recognition) tests. In all four experiments the explicit test revealed an advantage for retention when words in the study phase had been presented spaced as opposed to massed. In the implicit test however, three of the four experiments did not produce a significant advantage for spaced items over massed items. Further, other experiments that have reported significant differences between spaced and massed items for an implicit test are limited in magnitude. For example, Jacoby and Dallas (1981) reported significance for one experiment (4a), while the other failed to reach significance (4b).

A potential factor in not finding a spacing effect for the implicit test could be that words have been the stimuli commonly used in studies demonstrating parallel effects (Jacoby & Dallas, 1981; Perruchet, 1989). Whereas, the stimuli used in this experiment is fairly dissimilar from previous experiments. For example, previous experiments investigating this issue typically use identical presentations throughout. Usually, there are no stimuli that would confuse the subject (where it looked almost the same but not quite, with the exception of differences in typecase i.e. Scarborough et al., 1977). The fact that our experiment had such stimuli may have made it more difficult to demonstrate parallel effects. Actually, when the "different" items are eliminated the spaced ($\underline{M}$ = 83.68) and massed ($\underline{M}$ = 76.89) do differ in the predicted direction.

That recognition memory is affected by spacing and repetitions is well substantiated (see Hintzman, 1974 for a review) and supported here. Spaced items were better recognized than massed which were better recognized than items seen only once. Of interest as well was that "same" and "new" items were correctly identified equally well. However, identification of moves for which the performance had changed was poor ("different" items, $\underline{M}$ = 42.2%). This replication of the recognition data once again reinforces the statement of unintentional biases in gymnastic judging.

General Discussion

Bias is a common problem in judging sports where performance aesthetics is the outcome criterion (e.g., figure skating, gymnastics, dressage, synchronized swimming). Although judges strive to be objective in their assessment of performance, there are numerous influences that bias these judgements. Some of these influences are deliberate (i.e., political; Ansorge & Scheer, 1988) and may be unavoidable, while other biases may be less explicit. A common bias is a judge's expectation of performance from an athlete -- top-rank calibre athletes often are given higher marks for a performance that is equivalent to an athlete that is either less well known or of lower rank. This bias is a result of an influence on either the perception or decision making processes, and the judge may or may not be aware of it.

A memory-influenced bias that has been given much less attention is the effect of viewing a warm-up on a judge's perception of the competitive performance. This bias has not been considered important since, in theory, the judge is assumed to have a relatively abstract knowledge base upon which to assess performance. Being abstract, each performance would be judged on its own merit and the impact of one specific performance would not be expected to have a lasting impression. Thus, the effect of a single exposure to a move during the warm-up would not be expected to bias the assessment of that move during the competitive performance. Indeed, for events such as gymnastics and figure skating, judges are free to watch the warm-up that occurs just prior to competition as well as the pre-competition training that occurs days or weeks before.

Three experiments were reported that examined if the memory of a prior exposure(s) to a gymnastic move affected the later assessment of performance, and whether or not the judges were aware of the prior performance. A summary of the results for the three experiments are presented in

Figures 3 and 4. The data from the perceptual test phases of these experiments were pooled (using only the "new", "same" and "different" once seen items from Experiment 3) and analyzed using a 3 (experiment: 1, 2, and 3) x 2 (performance: perfect/error) x 3 (type: same/different/new) with repeated measures on the last two factors. Moves that were the same at study and test attained the highest level of accuracy ($\underline{M}$ = 76.2%). New moves were less accurate ($\underline{M}$= 72.2%). The lowest level of accuracy occurred for moves which had changed in performance from study to test ($\underline{M}$ = 68.4%). This main effect of item type was reliable at $\underline{p}$ = .001, $\underline{F}(2,100)$ = 7.32, $\underline{MSe}$ = 212.09. With certainty, prior processing of a gymnastic move affects later perceptual judgements.

A main effect for experiment approached conventional levels of significance ($\underline{p}$ = .055), $\underline{F}(2,50)$ = 3.03, $\underline{MSe}$ = 425.84. The absence of an interaction indicated that the influences of a prior processing of gymnastic moves were similar across experiments. In addition, a main effect for performance, $\underline{F}$ (1,50) = 15.31, $\underline{MSe}$ = 1070.68, was also found.

Insert Figure 3 about here

These differences in perceptual accuracy are not only of statistical significance but also of practical significance. While a difference of 7% does not appear to be a large effect, the past two Olympics clearly illustrate the potential ramifications. The individual standings in both Olympics are dramatic examples. The maximum attainable score for an individual gymnast is 80 points. In 1984, Mary Lou Retton, the first place medal winner scored 79.175, and Kathy Johnson, the tenth place gymnast scored 77.450. The 1988 Olympics also showed a close score between Elana Shushonova's (gold medallist) score of 79.675 and Brandy Johnson's (tenth place) score of 78.550. These scores of the gold medallists and the tenth place finishers were less than a 4% difference (International Gymnast, 1984, 1988).

By these comparisons, the impact of a 7% difference (or even a 5% difference, which was not statistically significant in Experiment 2) in judgement accuracy could have serious consequences on a gymnast's placing in a competition. For instance, if the warm-up is performed without error, then a duplicated performance during competition will be scored optimally. Furthermore, a later performance that is flawed will be scored better than it probably deserved. However, if the warm-up is flawed, a duplicated flawed performance during competition will be given a minimal score. And, a later errorless performance during competition will be scored lower than the performance deserved.

Figure 4 presents the recognition results of all three experiments -- a convincing picture of the judges poor level of awareness for recognition of "different" items as new. These recognition data were pooled (ignoring massed and spaced recognition data from experiment 3) and analyzed using a 3 (experiment: 1, 2, and 3) x 2 (performance: perfect/error) x 3 (type: same/different/new) ANOVA with repeated measures on the last two factors. Recognition of the "same" items (M = 61.8%) and "new" items (M = 22.5%) was accurate. However, correct identification of "different" items was poor (48%). This main effect of type was highly significant, $F(2,100) = 55.13$, $MSe = 718.03$. In addition, the overall recognition scores were significantly lower for Experiment 3 (M = 34.5%) than for Experiments 1 (M = 50.6%) and 2 (M = 46.3%). This difference could be attributed to the increased number of stimuli that subjects viewed in the study phase of Experiment 3. This difference was supported by a main effect for experiment, $F(2,50) = 5.64$, MSe = 1226.26. However, the absence of an interaction reveals that influences were similar across all three experiments.

Insert Figure 5 about here

Furthermore, Experiment 2 investigated if recognition of the "different" moves affected the perception judgements. Similar influences were noted for items correctly identified as new and items incorrectly identified as old. Thus, recognition of an item's change in performance did not appear to eliminate influences of prior processing. This evidence from the recognition phase of the experiment revealed that unintentional influences were a source of bias during the perceptual test phase. Thus, judges appeared to be unable to discount the influence of a prior exposure on their perception of a gymnastic move's performance.

The practical implications of the present study are evident for both the athlete and judge. Since the warm-up plays such an important role in determining their score, it is imperative that the gymnast enter into the warm-up activity with the same intensity and conviction to achieve as is brought to the competitive performance. Moreover, to improve objectivity of judges, efforts should be made to prevent the judges from seeing the warm-up.

Neisser (1982) argued that psychology has followed two routes in the study of memory. One route, is to discover basic mental mechanisms that can be revealed by well-controlled experiments. The other route, emphasizing the importance of ecological validity, is to understand the common examples of memory in ordinary human experience. In Neisser's view, the latter should be attributed more attention than it has received. Perhaps one strength of the present series of experiments is the convergence of these two separate routes -- the ideas of current importance in cognitive psychology (prior processing effects) applied to a real-world setting (judging gymnastics).

The findings in these experiments may be generalizable to sports other than gymnastics, where judges are free to watch an athlete's performance immediately prior to competition (i.e. dressage, figure skating, diving and synchronized swimming). In the same way that gymnastic judges are affected, prior episodes could easily affect the judges of other sports as well. Indeed, one often hears television commentators referring back to an athlete's performance during a warm-up. Is it reasonable to expect a judge to be any less cognizant?

The results of these experiments leaves some interesting questions regarding prior processing effects on judging gymnastics. Are these effects long-lasting? Could judges be influenced by the last competition in which he/she judged the same gymnast? Are these influences limited to the sport of gymnastics? Or more generally, limited to sports with a large subjective component? Both the theoretical and applied aspects of this research merit continued investigation.

References

Allard, F. A., & Starkes, J. L. (1980). Perception in sport: Volleyball. Journal of Sport Psychology, 2, 22-33.

Ansorge, C.J., & Scheer, J.K., Laub, J., & Howard, J. (1978). Bias in judging women's gymnastics induced by expectations of with-in team order. Research Quarterly, 49, 399-405.

Ansorge, C.J., & Scheer, J.K. (1988). International bias detected in judging gymnastic competition at the 1984 Olympic Games. Research Quarterly for Exercise and Sport, 59, 103-107.

Armstrong, C. W., & Hoffman, S. J. (1979). Effects of teaching experience, knowledge of performer competence, and knowledge of performance outcome on performance error identification. Research Quarterly, 50, 318-327.

Biscan, D. V., & Hoffman, S. J. (1976). Movement analysis as a generic ability of physical education teachers and students. Research Quarterly, 47, 161-163.

Bowles, N. L., & Glanzer, M. (1983). An analysis of interference in recognition memory. Memory and Cognition, 11, 307-315.

Cuddy, L. J., & Jacoby, L. L. (1982). When forgetting helps memory: An analysis of repetition effects. Journal of Verbal Learning and Verbal Behavior, 21, 451-467.

Eich, E. (1984). Memory for unattended events: Remembering with and without awareness. Memory and Cognition, 12, 105-111.

Feustal, T. C., Shiffrin, R. M., & Salasoo, A. (1980). Episodic and lexical contributions to the repetition effect in word identification. Journal of Experimental Psychology: General, 112, 309-346.

Fischoff, B. (1975). Hindsight = Foresight: The effect of outcome knowledge on judgement under uncertainty. Journal of Experimental Psychology: Human Perception and Performance, 1, 288-299.

Glaser, R., & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R, Glaser, & M. J. Farr (Eds.), The nature of expertise (pp. xv-xxviii). Hillsdale N. J.: Erlbaum.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. Journal of Verbal Learning and Verbal Behavior, 23, 553-568.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 501-518.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), Theories in cognitive psychology: The Loyola Symposium (pp. 77-99) Potomac, MD: Erlbaum.

Jacoby, L. L. (1978). On interpreting the effects of repetitions: Solving a problem versus remembering a solution. Journal of Verbal Learning and Verbal Behavior, 17, 649-667.

Jacoby, L.L. (1983a). Perceptual enhancement: Persistent effects of an experience. Journal of Experimental Psychology: Learning, Memory and Cognition, 9, 21-38.

Jacoby, L.L. (1983b). Remembering the data: Analyzing interactive processes in reading. Journal of Verbal Learning and Verbal Behavior, 22, 485-508.

Jacoby, L.L. (1988). Personal communication.

Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: Noise judgements. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 240-247.

Jacoby, L.L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General, 110, 306-340.

Jacoby, L. L., & Kelley, C. M. (1987). Unconscious influences of memory for a prior event. Personality and Social Psychology Bulletin, 13, 314-336.

Kolers, P. A. (1976). Reading a year later. Journal of Experimental Psychology: Human Learning and Memory, 2, 554-556.

Melton, A. W. (1967). Repetition and retrieval from memory. Science, 158, 532.

Murrel, G. A., & Morton, J. (1974). Word recognition and morphemic structure. Journal of Experimental Psychology: General, 102, 963-968.

Neisser, U. (1982). Memory observed. San Fransisco: Freeman.

Perruchet, P. (1989). The effect of spaced practice on explicit and implicit memory. British Journal of Psychology, 80, 113-130.

Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. Annual Review of Psychology, 39, 475-543.

Salmela, J.H. (1978). Gymnastic judging: A complex information processing task, or (who's putting one over on who?). International Gymnast, 20, 54-56, 62 -63.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Performance, 3, 1-17.

Scarborough, D. L., Gerard, L., & Cortese, C. (1979). Accessing lexical memory: The transfer of word repetition effects across task and modality. Memory and Cognition, 7, 3-12.

Underwood, B. J. (1957). Interference and forgetting. Psychological Review, 64, 49-60.

Witherspoon, D., & Allan, L. G. (1985). The effects of a prior presentation on temporal judgements in a perceptual identification task. Memory & Cognition, 13, 101-111.

Witherspoon, D., & Moscovitch, M. (1989). Stochastic independence between two implicit memory tasks. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 22-30.

Women's Results: XXIIIrd Olympiad. (1984, November). International Gymnast, 26, 80-85.

Women's Results: XXIVth Olympiad. (1988, December). International Gymnast, 30, 70-71.

Footnotes

1. Subjects were instructed to look only for breaks in the execution of a move as a result of lack of body control. Common form errors are toes not pointed, legs bent, arms bent, etc. The judges were asked not to look for other types of errors, such as technique and amplitude.

2. Stimuli were previewed by three gymnastic experts. Only those gymnastic moves that were agreed upon by all three previewers were used for the experiment.

3. Two factors were implemented to avoid potential ceiling effects. First, the duration of the video clip was brief (one second). Second, many of the form errors were not extremely noticeable, detection would occur only if the judge was looking in the right place at the right time.

4. For example, suppose a move could be divided into a numbered sequence from one to ten and the perfect performance video clip consisted of a segment of the move from number two through to number eight. The error performance video clip would also consist of the same segment (two through eight). In the "error" video clip a form error occurred at some point during the taped segment.

5. On three of the events (uneven bars, balance beam and floor exercise) the new moves were all completely different. However, the vault was an exception. The nature of the event severely limits the number of moves that could be used for stimuli. Therefore, to create enough stimuli for the new moves, the same vaulting move would be used to make two different video clips. That is, one video clip could be of

the first portion (preflight), whereas another video clip could use the end portion (postflight) (with no overlapping frames). The judges were forewarned of this and were told to call a move "old" only if it was identical to one in the study phase -- nothing whatsoever should differ from the original.

Table 1.

Accuracy of Form Error Judgements During the Perceptual Test Phase (percent correct) in Experiment 1.

	STUDY-TEST TYPE			
TEST PERFORMANCE	Same	Different	New	Mean
Error	68.8	64.2	66.2	66.4
Perfect	88.5	79.8	84.7	84.3
Mean	78.7	72.1	75.4	

Table 2.

Scores of Recognition Phase Judgements (probability of judging an item as old) in Experiment 1.

TEST PERFORMANCE	STUDY-TEST TYPE Same	Different	New	Mean
Error	69.8	47.2	28.4	48.5
Perfect	68.8	50.0	28.1	49.0
Mean	69.3	48.6	28.3	

Table 3

Accuracy of Form Error Judgements During the Study Phase (percent correct) in Experiment 3.

REPETITION TYPE

PERFORMANCE	Once	Massed First	Massed Second	Spaced First	Spaced Second
Error	64.3	73.3	72.8	65.6	77.3
Perfect	81.8	75.0	81.2	81.5	80.9
Mean	73.0	74.2	77.0	73.5	79.1

Figure Captions

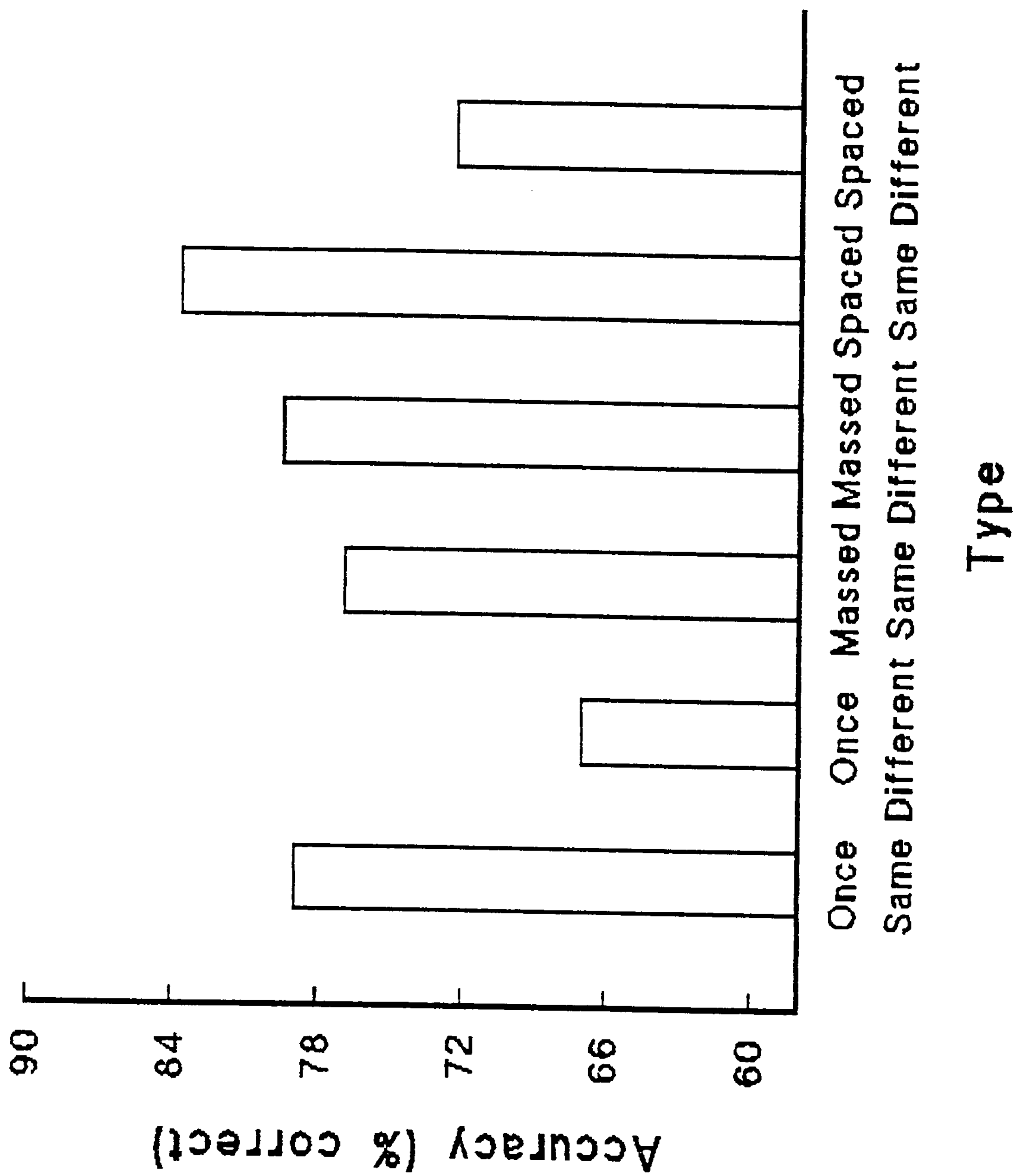
Figure 1. Accuracy of Form Error Judgements for the Perceptual Test Phase (percent correct) in Experiment 3.

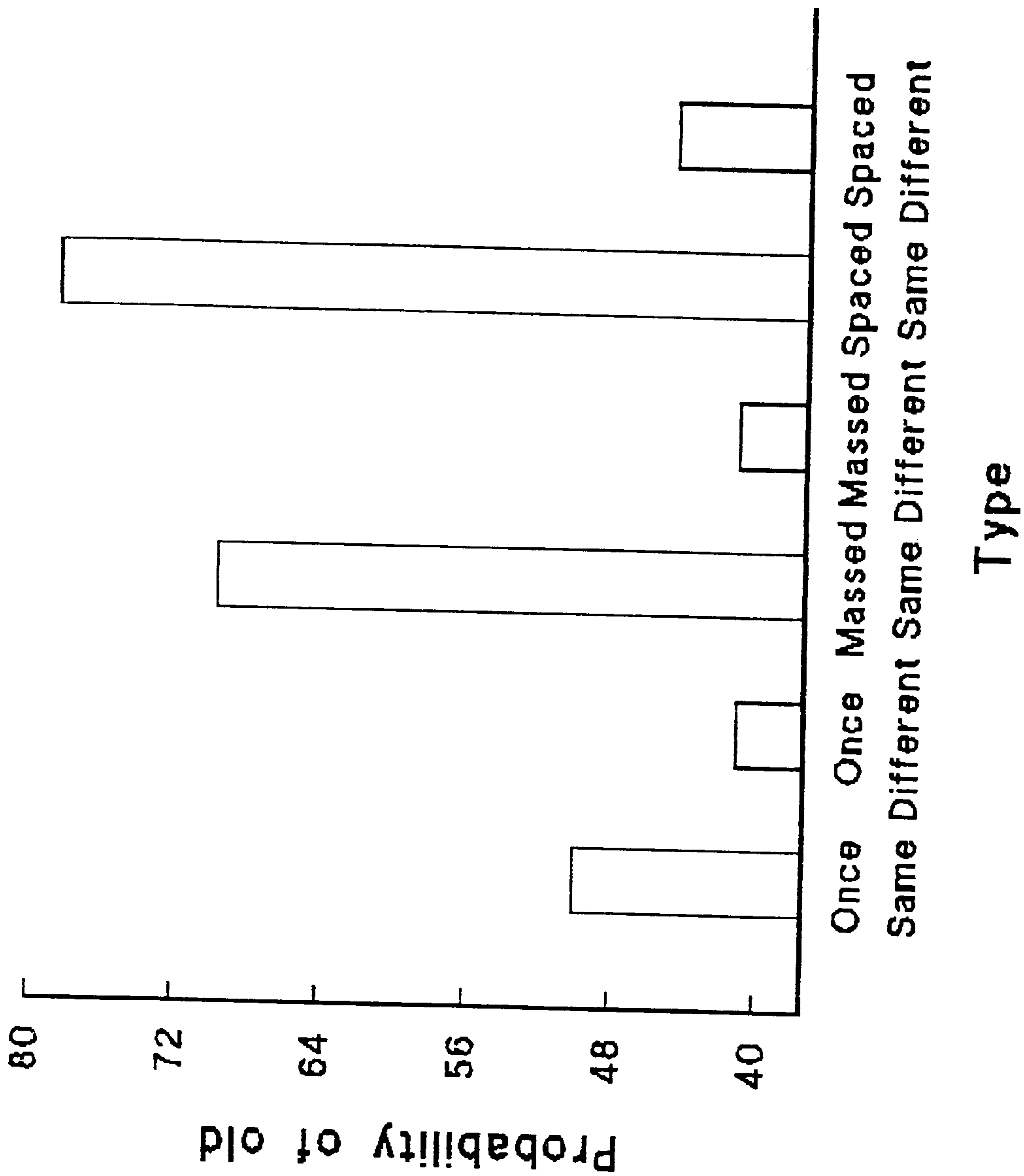
Figure 2. Scores of Recognition Judgements (probability of judging an item old) in Experiment 3.

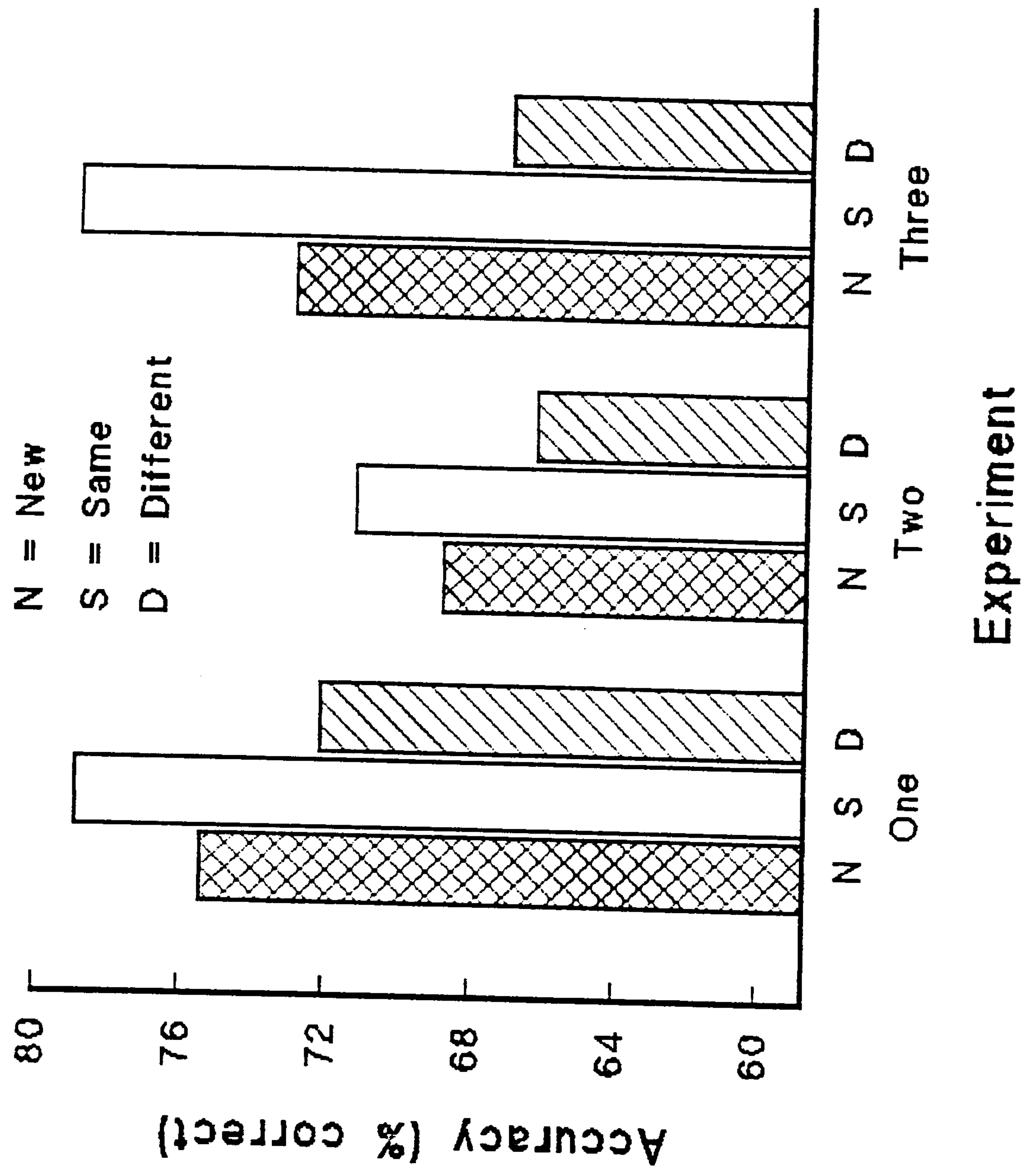
Figure 3. Accuracy of Form Error Judgements (percent correct, Perceptual Test Phases) for Experiments 1, 2, and 3.

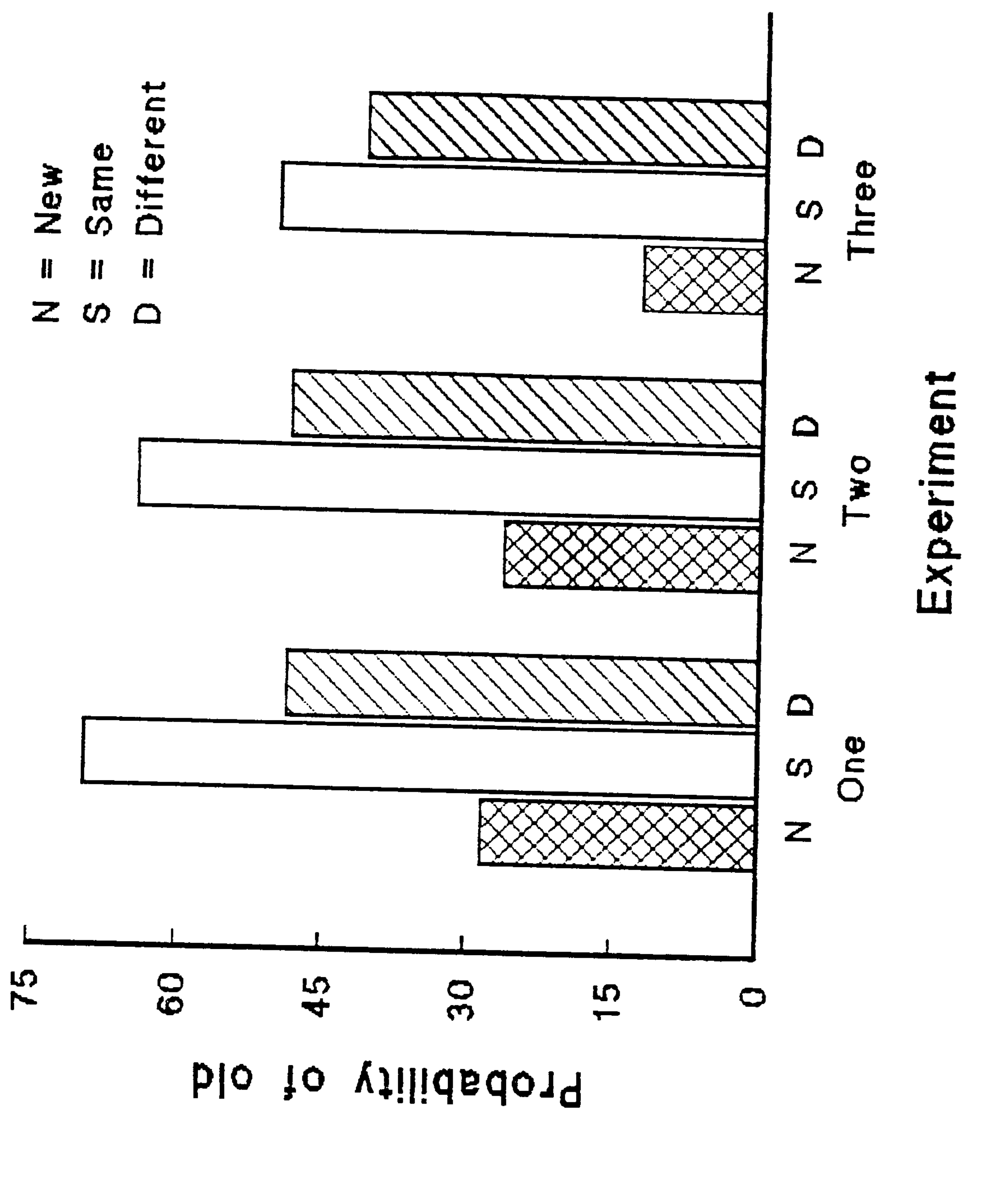
Figure 4. Scores of Recognition Judgements

(probability of judging an item old) for Experiments 1, 2, and 3.


Accuracy (% correct)
90
84
78
72
66
60
Once Same
Once Different
Massed Same
Massed Different
Spaced Same
Spaced Different
Type


Probability of old
80
72
64
56
48
40
Once Same
Once Different
Massed Same
Massed Different
Spaced Same
Spaced Different
Type


Accuracy (% correct)
80
76
72
68
64
60
N = New
S = Same
D = Different
N S D
One
N S D
Two
N S D
Three
Experiment


N = New
S = Same
D = Different
Probability of old
0
15
30
45
60
75
N S D
One
N S D
Two
N S D
Three
Experiment

APPENDICES

APPENDIX A
ANOVA Tables

Table 4

Study phase accuracy ANOVA for Experiment 1

Source	MS	df	F	
Group	483.870	1	7.54	*
Error	64.155	22		
Performance	2363.213	1	25.287	*
Grp x Perf	552.164	1	5.908	*
Error	93.454	22		

Table 5

Perceptual test phase accuracy ANOVA for Experiment 1

Source	MS	df	F	
Group	548.731	1	2.906	
Error	188.858	22		
Performance	11572.381	1	20.316	*
Grp x Perf	3663.274	1	6.431	*
Error	569.615	22		
Type	534.841	2	3.909	*
Grp x Type	14.740	2	.108	
Error	136.811	44		
Perf x Type	53.527	2	.248	
Grp x Perf X Type	272.831	2	1.265	
Error	215.733	44		

Table 6

Recognition test phase accuracy Anova for Experiment 1

Source	MS	df	F	
Group	435.507	1	.308	
Error	1419.883	22		
Performance	8.507	1	.018	
Grp x Perf	1116.673	1	2.347	
Error	475.802	22		
Type	20172.563	2	26.386	*
Grp x Type	316.840	2	.414	
Error	764.527	44		
Perf x Type	49.173	2	.103	
Grp x Perf x Type	743.257	2	1.559	
Error	476.723	44		

Table 7

Study phase accuracy ANOVA for Experiment 2

Source	MS	df	F
Performance	300.200	1	1.566
Error	191.625	14	

Table 8

Perceptual test accuracy ANOVA for Experiment 2

Source	MS	df	F
Performance	1876.899	1	1.057
Error	1775.419	14	
Type	186.507	2	1.249
Error	149.268	28	
Perf x Type	228.784	2	1.208
Error	189.389	28	

Table 9

Recognition test accuracy ANOVA for Experiment 2

Source	MS	df	F
Performance	1.849	1	.010
Error	181.300	14	
Type	10858.567	2	34.442 *
Error	315.272	28	
Perf x Type	204.379	2	1.528
Error	133.795	28	

Table 10

Test phase VRT ANOVA for Experiment 2

Source	MS	df	F
Performance	.037	1	.934
Error	.040	10	
Type	.013	2	.218
Error	.059	20	
Perf x Type	.003	2	.112
Error	.034	20	

Table 11

Study phase accuracy ANOVA for Experiment 3

Source	MS	df	F
Performance	3095.420	1	2.043
Error	1514.609	13	
Type	188.004	4	.992
Error	189.514	52	
Perf x Type	355.802	4	1.888
Error	188.522	52	

Table 12

Perceptual test phase accuracy ANOVA for Experiment 3

Source	MS	df	F	
Performance	3714.881	1	2.832	
Error	1311.893	13		
Repetition	495.792	2	1.165	
Error	425.477	26		
Perf x Rep	1457.327	2	2.808	
Error	518.898	26		
Type	1988.595	1	4.310	*
Error	461.351	13		
Perf x Type	152.381	1	.440	
Error	346.188	13		
Rep X Type	939.292	2	3.210	*
Error	292.605	26		
Perf x Rep x Type	569.042	2	.883	
Error	644.599	26		

Table 13

Recognition test phase accuracy ANOVA for Experiment 3

Source	MS	df	F	
Performance	952.381	1	.543	
Error	1753.663	13		
Repetition	3675.595	2	3.306	*
Error	1111.492	26		
Perf x Rep	1443.452	2	.818	
Error	1763.965	26		
Type	23809.524	1	20.717	*
Error	1149.267	13		
Perf x Type	59.523	1	.027	
Error	2142.857	13		
Rep x Type	2425.595	2	1.860	
Error	1303.800	26		
Perf x Rep x Type	2514.880	2	4.493	*
Error	559.753	26		

Table 14

Perceptual test phase VRT ANOVA for Experiment 3

Source	MS	df	F
Performance	.159	1	1.253
Error	.127	13	
Repetition	.033	2	.335
Error	.098	26	
Perf x Rep	.018	2	.204
Error	.089	26	
Type	.356	1	1.978
Error	.179	13	
Perf x Type	.009	1	.038
Error	.249	13	
Rep x Type	.031	2	.244
Error	.129	26	
Perf x Rep x Type	.020	2	.191
Error	.105	26	

Table 15

Recognition test phase VRT ANOVA for Experiment 3

Source	MS	df	F
Performance	.063	1	.407
Error	.154	13	
Repetition	.062	2	.419
Error	.147	26	
Perf x Rep	.260	2	1.274
Error	.204	26	
Type	.788	1	5.408 *
Error	.145	13	
Perf x Type	.0006	1	4.218
Error	.153	13	
Rep x Type	.018	2	.073
Error	.253	26	
Perf x Rep x Type	.163	2	.824
Error	.198	26	

Table 16

Perceptual test phase accuracy ANOVA for pooled data of Experiments 1,2, and 3

Source	MS	df	F	
Experiment	1292.63	2	3.035	*
Error	425.84	50		
Performance	16388.51	1	15.307	*
Error	1070.68	50		
Exp x Perf	592.30	2	.553	
Error	1070.68	50		
Type	1543.42	2	7.278	*
Error	212.09	100		
Exp x Type	97.48	4	.460	
Error	212.09	100		
Perf x Type	221.56	2	.823	
Error	269.25	100		
Exp x Perf x Type	465.43	4	1.73	
Error	269.25	100		

Table 17

Recognition test phase scores ANOVA for pooled data of Experiments 1, 2, and 3

Source	MS	df	F	
Experiment	6918.21	2	5.642	*
Error	1226.26	50		
Performance	101.67	1	.102	
Error	994.99	50		
Exp x Perf	814.56	2	.819	
Error	995.00	50		
Type	39588.07	2	55.134	*
Error	718.02	100		
Exp x Type	282.62	4	.394	
Error	718.03	100		
Perf x Type	1206.28	2	2.191	
Error	550.59	100		
Exp x Perf x Type	984.34	4	1.788	
Error		100		

APPENDIX B

Cell Means

Table 18

Study phase cell means for Experiment 1

Group	Performance	% Correct
Expert	Perfect	84.93
Expert	Error	77.67
Novice	Perfect	85.36
Novice	Error	64.54

Table 19

Perceptual test phase cell means for Experiment 1

Group	Performance	Type	% Correct
Expert	Perfect	New	83.03
Expert	Perfect	Same	86.46
Expert	Perfect	Different	74.25
Expert	Error	New	70.53
Expert	Error	Same	75.53
Expert	Error	Different	73.95
Novice	Perfect	New	86.31
Novice	Perfect	Same	90.63
Novice	Perfect	Different	85.36
Novice	Error	New	61.90
Novice	Error	Same	61.89
Novice	Error	Different	54.45

Table 20

Recognition test phase cell means for Experiment 1

Group	Performance	Type	% Probability of Old
Expert	Perfect	New	18.75
Expert	Perfect	Same	66.67
Expert	Perfect	Different	47.92
Expert	Error	New	31.92
Expert	Error	Same	74.33
Expert	Error	Different	42.33
Novice	Perfect	New	37.50
Novice	Perfect	Same	70.92
Novice	Perfect	Different	52.08
Novice	Error	New	25.00
Novice	Error	Same	65.25
Novice	Error	Different	52.08

Table 21

Study phase cell means for Experiment 2

Performance	% Correct
Perfect	83.03
Error	76.70

Table 22

Perceptual test phase cell means for Experiment 2

Performance	Type	% Correct
Perfect	New	75.63
Perfect	Same	73.03
Perfect	Different	70.67
Error	New	60.87
Error	Same	69.31
Error	Different	61.75

Table 23

<u>Recognition test phase cell means for Experiment 2</u>

Performance	Type	% Correct
Perfect	New	29.01
Perfect	Same	64.21
Perfect	Different	46.03
Error	New	23.47
Error	Same	63.98
Error	Different	50.94

Table 24

<u>VRT cell means for Experiment 2</u>

Performance	Type	Time to respond
Perfect	New	2930
Perfect	Same	2890
Perfect	Different	2940
Error	New	2850
Error	Same	2860
Error	Different	2910

Table 25

Study phase cell means for Experiment 3

Performance	Repetition	% Correct
Perfect	Once seen	81.79
Perfect	First Presentation spaced	81.50
Perfect	Second presentation spaced	80.86
Perfect	First presentation massed	75.01
Perfect	Second presentation massed	81.22
Error	Once seen	64.27
Error	First presentation spaced	65.60
Error	Second presentation spaced	77.34
Error	First presentation massed	73.34
Error	Second presentation massed	72.83

Table 26

Perceptual test phase cell means for Experiment 3

Performance	Repetition	Type	% Correct
Perfect	Once seen	Same	85.14
Perfect	Once seen	Different	81.57
Perfect	Twice seen spaced	Same	83.36
Perfect	Twice seen spaced	Different	73.86
Perfect	Twice seen massed	Same	82.21
Perfect	Twice seen massed	Different	80.36
Error	Once seen	Same	72.64
Error	Once seen	Different	52.43
Error	Twice seen spaced	Same	84.00
Error	Twice seen spaced	Different	70.86
Error	Twice seen massed	Same	71.57
Error	Twice seen massed	Different	78.57

Table 27

Recognition test phase cell means for Experiment 3

Performance	Repetition	Type	Probability of old
Perfect	Once seen	Same	53.57
Perfect	Once seen	Different	53.57
Perfect	Twice seen spaced	Same	75.00
Perfect	Twice seen spaced	Different	50.00
Perfect	Twice seen massed	Same	75.00
Perfect	Twice seen massed	Different	32.14
Error	Once seen	Same	46.43
Error	Once seen	Different	28.57
Error	Twice seen spaced	Same	82.14
Error	Twice seen spaced	Different	39.29
Error	Twice seen massed	Same	64.29
Error	Twice seen massed	Different	50.00

Table 28

Perceptual test phase VRT cell means for Experiment 3

Performance	Repetition	Type	Response Time
Perfect	Once seen	Same	2360
Perfect	Once seen	Different	2470
Perfect	Twice seen spaced	Same	2290
Perfect	Twice seen spaced	Different	2420
Perfect	Twice seen massed	Same	2370
Perfect	Twice seen massed	Different	2460
Error	Once seen	Same	2370
Error	Once seen	Different	2360
Error	Twice seen spaced	Same	2240
Error	Twice seen spaced	Different	2410
Error	Twice seen massed	Same	2280
Error	Twice seen massed	Different	2350

Table 29

Recognition test phase VRT cell means for Experiment 3

Performance	Repetition	Type	Response Time
Perfect	Once seen	Same	2710
Perfect	Once seen	Different	2860
Perfect	Twice seen spaced	Same	2740
Perfect	Twice seen spaced	Different	2750
Perfect	Twice seen massed	Same	2800
Perfect	Twice seen massed	Different	3060
Error	Once seen	Same	2730
Error	Once seen	Different	2930
Error	Twice seen spaced	Same	2680
Error	Twice seen spaced	Different	2880
Error	Twice seen massed	Same	2720
Error	Twice seen massed	Different	2740