

Retrieval plus scanning: Does it occur?

SUSAN BRANNELLY, GERALD TEHAN, and MICHAEL S. HUMPHREYS
University of Queensland, St. Lucia, Queensland, Australia

In the two experiments reported here, we tested the retrieval-plus-scan model of delayed probe recognition by adding a second probe 2 sec after subjects had responded to the first probe. According to this model, the list items should still be in consciousness from the first probe at the time of the second probe. Consequently, on tests in which the first probe had been tested immediately, we expected to find the same pattern of performance on both first and second probes. On tests in which the first probe came after a filled delay, we expected the first-probe data to show the effects of retrieval and proactive interference. These effects should not be present on the second probe. The results of the first-probe data in both experiments were consistent with the retrieval-plus-scan model. The second-probe data did not fit the model, however. In the first experiment, the second probe on what had been delayed first-probe trials still produced intercept differences, indicating the presence of retrieval. In the second experiment, the effects of proactive interference were still present on the second probe. It is suggested that these results not only invalidate the retrieval-plus-scan model for delayed probe recognition but that they, along with other results, cast doubt on a scanning operation in immediate probe recognition.

Recall at short intervals seems to be qualitatively different from recall at longer retention intervals (Craik & Levy, 1976). The initial explanations for this difference proposed that different memory structures underlay performance at these different retention intervals. Primary memory (PM) in the form of a short-term store was seen to underlie performance on short-term memory tasks, and secondary memory (SM) in the form of a long-term store was seen to be responsible for performance on long-term memory tasks. The evidence for this distinction was initially quite strong, but over the years support for several of the central tenets of these theories was shown to be lacking, with the result that the evidential basis for making the distinction between PM and SM has been slowly eroding (Crowder, 1982).

Recently, however, the work of Wickens and his associates (Wickens, Moody, & Dow, 1981; Wickens, Moody, & Vidulich, 1985) has suggested a methodology that appears to operationalize the distinction between PM and SM. Their model of PM is one of a number that equates the contents of PM with what is in consciousness (Craik & Jacoby, 1975; James, 1890). Their operationalization of PM and SM is based upon differential performance on immediate and delayed trials of the Sternberg recognition paradigm. In Wickens et al.'s version of this task, subjects are shown a small number of items (either two or four items in the memory set) and then are asked to recognize a probe after either a brief (2-sec), unfilled

retention interval (immediate test) or a 12-sec interval (delayed test) filled with some form of distracting activity such as counting backward in threes from a random three-digit number. The effects of proactive interference (PI) also are incorporated into the task by observing recognition performance over a number of trials using materials from the same taxonomic category. Specifically, instances from the same category are used over three trials, another set of materials is used for the next three trials, and so on. This procedure sets up a continuous release from a PI situation in that on the first trial of any given set of materials, PI will be at a minimum. PI then builds up over the next trials so that it is at a maximum by the third trial.

In all of their experiments using this methodology, effects of memory-set size were found. It took subjects longer to recognize a positive probe, or to reject a negative probe, if there were four items in the memory set than if there were only two. Moreover, in all of their conditions, the scanning rates (i.e., the differences in reaction time between the 2- and 4-item memory sets) were equivalent. The only differences lay in the intercept values for each condition.

These intercept differences are important for the present paper. On immediate trials, Wickens et al. (1981) observed no difference in latency between low- and high-interference trials. However, an intercept difference emerged when the tests were delayed. Subjects responded to delayed low-interference trials 95 msec slower than they responded to immediate trials. In addition, the effects of PI emerged: latencies for high-interference delayed trials were reliably slower (about 40 msec) than latencies for low-interference delayed trials.

Wickens et al.'s explanation for this pattern of results is that items are in consciousness on immediate trials. When the probe item is presented, it is matched with each item

This research was supported by a grant from the Australian Research Council to the third author. Experiment 1 and the first cohort of Experiment 2 are taken from the first author's honors thesis at the University of Queensland. Reprint requests should be addressed to Michael S. Humphreys, Department of Psychology, University of Queensland, St. Lucia, Queensland 4067, Australia.

in consciousness in a serial exhaustive search (Sternberg, 1969), and since it takes longer to search large sets than small sets, latencies differ for different set sizes. In addition, since the items are already in consciousness, they do not have to be retrieved. As consequence, immediate trials are free from the effects of PI, since the locus of PI effects is generally accepted to be at the retrieval stage of processing (Watkins & Watkins, 1975). It is on the basis of the lack of retrieval and PI effects that Wickens et al. concluded that immediate trials are an operationalization of PM.

Wickens et al.'s explanation of equal scanning rates but different intercepts for the high and low PI conditions was that either the list as a whole or a pointer to the list was retrieved from SM and made active in PM. This retrieval took time and was subject to the effects of PI. Once the list had been retrieved (brought back into consciousness), however, it was searched in the same way as on immediate trials.

Halford, Maybery, and Bain (1988) have provided additional evidence that seems consistent with some aspects of Wickens et al.'s analysis. On the basis of the general assumption in the literature that PM is of limited capacity, Halford et al. argued that if one increases the set size beyond the capacity of PM, some of the items will have to be retrieved from SM. Thus, even on an immediate test, the effects of retrieval and PI should become evident on large set sizes. Using set sizes of 4, 6, 8, and 10 items, they replicated Wickens et al.'s finding of no PI effects for a set size of four, but they did find PI effects for all the other set sizes used. Halford et al. then tested the hypothesis that PM capacity increases with age by testing a group of 8- and 9-year-olds. They found that no PI was observed with a set size of two but that PI was present with a set size of four. These results are compatible with the notion that items from SM have to be retrieved (but see Halford et al. for some alternative explanations).

The experiments described in this paper were designed to test a particular aspect of the Wickens et al. model, though they may have implications for the remaining aspects of the model as well. As mentioned earlier, performance on a delayed task has been explained in terms of the list being retrieved, with the retrieval being subject to PI. Once the list has been retrieved (brought back into consciousness), the list is scanned in an identical fashion as in the PM case. The question we posed is, what would happen if a second probe was presented shortly after the subject had made his or her decision about the first probe? On the basis of the Wickens et al. model, we expected no retrieval or PI effects with the second probes. On an immediate test, the list items are assumed to be in consciousness prior to the presentation of the first probe. The serial scan, in which the probe is compared to each item in the list, presumably serves to maintain the list items in PM (i.e., it serves as a rehearsal task). Thus, when the second probe is presented, the list items should still be in PM and the reaction time to the second probe should be independent of the level of PI. With a delayed test, the initial situation is different but the second-

probe results should be similar to the results found on an immediate test. Prior to the first probe, the items are presumed to be in SM. When the probe is presented, the list is assumed to be retrieved. Following this retrieval, the list items are assumed to be scanned just as if they had always been in PM. Now when a second probe is presented, there is presumably no need to retrieve the list items. The second scan on this list should be identical to the second scan on an immediate test. Thus, on delayed tests, we would expect to see the effects of retrieval and PI on the first probe, but no retrieval and PI effects on the second probe. The two experiments which follow were set out to examine this prediction.

In the first experiment, we examined what Wickens et al. call *uncomplicated retrieval*. We examined latencies for first and second probes on immediate and delayed tests, in which the items studied were uncategorized and were used only once in the experiment (except when the item was a positive probe). We predicted that for the first probe, we would replicate Wickens et al.'s finding of identical scanning rates but different intercepts for the immediate and delayed conditions. For the second probe, we expected to find no difference in either scanning rates or intercepts between immediate and delayed trials since it was argued that in both cases the items were already in PM.

EXPERIMENT 1

Method

Subjects

Forty, first-year psychology students from the University of Queensland served as subjects in the experiment. The subjects were tested individually and all received course credit for their participation. The data of 1 of the subjects were eliminated from all analyses because a computer error resulted in only a small portion of the data being recorded.

Materials and Apparatus

The experimental setting was a sound-attenuated room. All stimuli were controlled and subject performance was recorded by an Amiga 1000 computer. The word stimuli for the practice trials, the *input trials*, and the *word-recognition probes* consisted of 344 high-frequency, high-concrete, and high-imagery nouns selected from either the Toronto Word Pool (Friendly, Franklin, Hoffman, & Rubin, 1982) or the Paivio, Yuille, and Madigan (1968) norms.

The "mouse" buttons on the Amiga served as response switches, one button being marked "yes," the other "no." No attempt was made to control the way in which subjects used the mouse to respond.

Procedure and Design

The task was similar to that employed by Wickens et al. (1981), with the exception that a second probe was presented after the first. Basically, the subjects performed two versions of the varied-set Sternberg (1969) paradigm, the versions differing only in the retention interval involved: a 2-sec unfilled interval in what Wickens et al. have termed the PM case, and a 12-sec filled-retention interval in the SM case. Each trial started with a brief "ready" signal, followed by a memory set of either two or four words which was presented simultaneously for 3 sec on the screen in a vertical array. The subjects were instructed to read the words aloud. On PM trials, after a 2-sec unfilled retention interval following the presentation of the material, a probe item appeared for 2 sec. The subject's task

was to decide whether or not the probe item was a member of the memory set and then to respond as quickly as possible by pressing one of the buttons on the mouse. After the response to the first probe, the screen went blank for 2 sec and then a second probe appeared for 2 sec. Again, the subject had to decide if this second probe was also a member of the memory set. In the SM trials, a random three-digit number appeared on the screen for 12 sec. The subjects were requested to say the number aloud and then to count backward in threes from that number. After 12 sec of counting, the screen cleared for 2 sec and then the first probe appeared for 2 sec. After the subjects had responded to the first probe, the screen cleared for 2 sec more and then the second probe appeared for 2 sec. The subjects then responded to this second probe. A 4-sec intertrial interval separated all trials. The 2-sec unfilled interval prior to the first probe was included both because Wickens et al. (1981) had used it and because we wanted to provide the same amount of preparation time prior to the first and second probes. As we shall see later, however, this may not have been successful.

The experiment consisted of 92 trials, which were presented in three blocks. The first block was a set of 16 practice trials followed by two experimental blocks of 38 trials, the first two trials of each experimental block also being practice trials. Each block was separated by a 3-min rest break. Of the 72 trials for which data were recorded, half the trials were immediate (PM) trials and half were delayed (SM) trials. Within these groups, half the trials consisted of two-item memory sets and half had four-item memory sets.

Each trial involved a subject making two recognition responses, one to the first probe and one to the second probe. There were four combinations of probes used in the experiment. These were Old Old (OO), Old New (ON), New Old (NO), and New New (NN), the first of each combination describing the first probe, the second referring to the second probe. "Old" represents a positive probe, that is, a word that was present in the memory set, requiring a "yes" response from the subjects. "New" words were words that had not been studied in the memory set, and thus required a "no" response from the subjects.

For the practice trials, all four probe combinations were equally distributed. This was not so for the experimental trials, since our primary interest was with recognition of Old words. Therefore, of the 72 experimental trials, 32 had the probe combination of OO, 16 of ON, 16 of NO, and 8 of NN; half of these trials appeared in the first block and half in the second. The choice of Old probes was randomly selected without replacement on each trial.

The assignment of items to trials was randomized for each subject, as was the order of trials within each block, with the constraint that not more than four repetitions of the same probe condition or three of the same-size memory set could occur. None of the word stimuli were ever seen more than once by any subject, except in the case of a positive probe.

Results

Two measures were used in the subsequent analyses: the median RTs of correct responses to the different types of probes, and the accuracy of the subjects' responses to the different types of probes. In all analyses, the data were submitted to 2×2 repeated-measures ANOVAs, with retention interval and memory-set size being the two factors. An alpha level of .05 was used to determine statistical reliability.

First-Probe Analyses

Although performance on the first probe trials was not the principle focus of this experiment, analysis of the first-

probe data should replicate Wickens et al.'s (1981) data. Since the subjects could not distinguish, at the time of responding, between the two conditions that had positive probes as the first probe (OO, ON), the data have been collapsed across these two sets of trials. Likewise, the data for the two conditions that had negative probes as the first probe (NO, NN) were collapsed. The means of median latency for the positive and negative first probes on error-free trials are presented in Figure 1.

Positive probes. The analysis of positive-probe latency revealed reliable effects for retention interval [$F(1,38) = 4.17$, $MS_e = 16,028$] and set size [$F(1,39) = 10.41$, $MS_e = 7,515$] (see Figure 1). The interaction between retention interval and set size was not significant.

The subjects correctly identified positive probes more often on the immediate test than on the delayed test [$F(1,38) = 20.77$, $MS_e = 0.004$], but this was qualified by an interaction with set size [$F(1,39) = 12.47$, $MS_e = 0.004$]. Performance was very accurate on immediate tests (98% and 99% accuracy for set sizes of two and four words, respectively); however, although performance was still accurate on a delayed test using a set size of two (97% accuracy), errors increased when a set size of four was probed after a delay (82% accuracy).

Negative probes. Analysis of negative-probe latency revealed no reliable results for retention interval, set size, or the interaction of retention interval with set size. The differences between means, although in the right direction, were not as pronounced as with the positive probes. In addition, performance on the negative probes seemed to be more variable than performance on the positive probes.

The accuracy data also appear to be equivalent across the various conditions (97% accuracy on immediate tests and 98% accuracy on delayed tests). There were no reliable main effects, nor a reliable interaction.

As can be seen in Figure 1, the latency data for the positive probes conform to the predictions of Wickens et al. (1981). On the immediate test, we found the traditional set-size effect associated with the Sternberg paradigm. We replicated the Wickens et al. data on the delayed test in that we obtained a difference in intercept but not in slope between the immediate test and the delayed test.

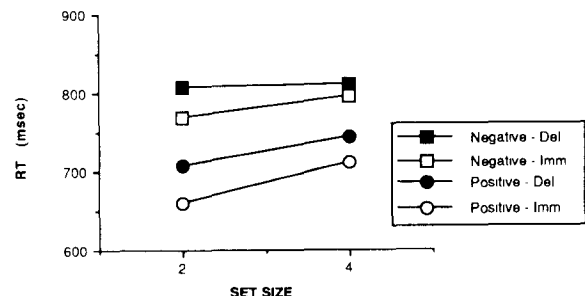


Figure 1. First-probe data for Experiment 1. Reaction time as a function of memory-set size, retention interval, and nature of the probe.

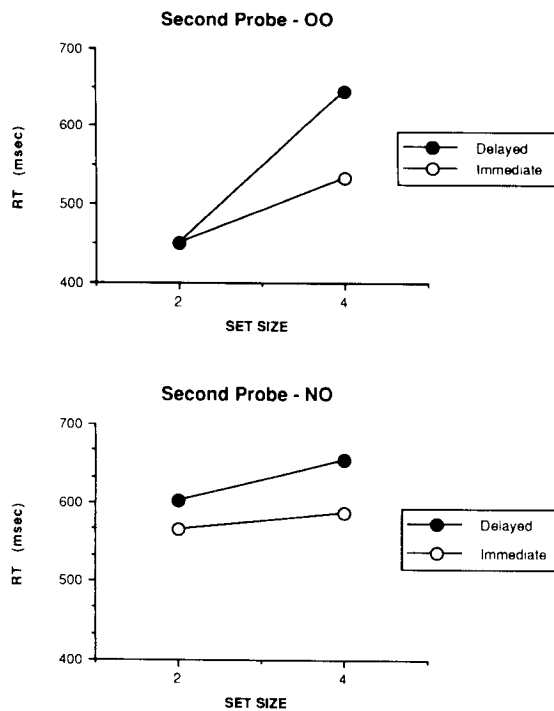


Figure 2. Second-probe data for the OO probe combination (top) and the NO probe combination (bottom) in Experiment 1. Reaction time as a function of memory-set size, retention interval, and nature of the probe.

Second-Probe Analyses

Positive probes. The means of median latency for the second probe (positive probes only) on error-free trials are presented in Figure 2. The data are presented separately for Probe Combinations OO and NO because of the obvious differences between these two conditions. That is, with Probe Combination OO, there was a difference between the immediate and delayed conditions with a set size of four but not with a set size of two. With Probe Combination NO, however, there was a difference between the immediate and delayed conditions with both set sizes.

The planned analysis was collapsed over Probe Combinations OO and NO. Reliable effects were found for set size [$F(1,38) = 57.68, MS_e = 7,097$] and retention interval [$F(1,38) = 18.98, MS_e = 6,097$]. In addition, the interaction between set size and retention interval was significant [$F(1,38) = 8.33, MS_e = 7,491$]. The accuracy data mirrored the latency data with main effects for set size and retention interval [$F(1,38) = 29.56, MS_e = 0.003$, and $F(1,38) = 34.97, MS_e = 0.004$, respectively]. The interaction between set size and retention interval was also significant [$F(1,38) = 17.84, MS_e = 0.003$]. Performance was at the 98% level on immediate tests; on delayed tests, however, although accuracy was still good with set sizes of two (97%), with set sizes of four, accuracy fell to the 88% level.

Negative probes. The pattern of RT performance on the negative second probes (Probe Combinations ON and

NN were combined and latencies were calculated as for the positive probes) was similar to that of the positive probes. The main effects of retention interval [$F(1,38) = 6.87, MS_e = 3,346$] and set size [$F(1,38) = 24.65, MS_e = 4,871$] were reliable, as was the interaction between retention interval and set size [$F(1,38) = 4.19, MS_e = 4,156$]. The median latencies were 770 and 796 msec for the immediate tests with set sizes of two and four, respectively; they were 808 and 812 msec for the delayed tests with set sizes of two and four, respectively. As was the case with the negative first probes, accuracy on the negative second probes was equivalent for all conditions at the 98% level.

These results provide only partial support for the notion that the list items were in PM for the second probe. The data for the two-word set size are consistent with the theory, but the data for the four-word set size are not. To examine these effects further, we looked, on a post hoc basis, at the two positive second probes (OO and NO) individually to see if the same pattern of performance held for both.

For the OO trials, the effects of retention interval and set size were reliable [$F(1,38) = 18.17, MS_e = 6,742$, and $F(1,38) = 106.48, MS_e = 7037$, respectively]. In addition, the interaction between retention interval and set size was significant [$F(1,38) = 14.67, MS_e = 8,244$].

As can be seen in the top panel of Figure 2, there was no difference on the second probe between the immediate and the delayed test when a set size of two was being used. This is the pattern of results one would expect if the set was in consciousness after the first probe had been processed and was scanned a second time. However, when the set size was four, there was a reliable difference between the immediate and the delayed test. This result is at odds with the notion that the list is in consciousness the second time it is probed.

Analysis of the NO probes produced only one reliable effect—that of retention interval [$F(1,38) = 5.83, MS_e = 17,733$]. The data are presented in the bottom panel of Figure 2. Although the means are based on only half the number of observations as for the OO combination, and are thus not as reliable as one would like, it would seem that the effects of retention interval have been replicated at the four-word set size, but that at the two-word set size, the second probe on the delayed test produced longer latencies than did the corresponding probe on the immediate test. Thus, we seem to have failed to replicate the two-word set-size effect found with the OO combination.

One aspect of the data that came as somewhat of a surprise was the general decrease in latency on the second probe. We had expected that all second-probe responses would be as fast as the responses to the first probe. To check the reliability of this observation, the four-word set-size data were subjected to a $2 \times 2 \times 2$ repeated-measures ANOVA, with probe number, test delay, and whether the probe was old or new being the factors. The reason that only four-word set-size data were used was that we believe that on trials with a set size of two, the subjects may

have used a recall strategy. Our reasons for this conclusion are presented more fully in the discussion.

Responses to the second probes were reliably faster than responses to first probes [$F(1,38) = 53.61$, $MS_e = 27,716$], longer latencies were found on delayed tests than on immediate tests [$F(1,38) = 8.38$, $MS_e = 21,020$], and the subjects responded more quickly to old probes than to new probes [$F(1,38) = 39.26$, $MS_e = 10,193$]. There was also a significant interaction between probe number and probe delay [$F(1,38) = 5.49$, $MS_e = 7,151$]. The difference in latency between the first and second probes was more pronounced on the immediate test (160 msec) than on the delayed test (116 msec).

Discussion

The first-probe data are consistent with the Wickens et al. model. Latencies were faster with a set size of two than with a set size of four. When the test was delayed, we found equivalent scanning rates in the immediate test condition, but we found a difference in intercept. Wickens et al. interpreted this pattern in terms of a serial exhaustive scan of the contents of PM in each case, the only difference being that in the delayed test, the items have to be retrieved from SM and made active in PM. The first-probe data therefore provide us with a benchmark that suggests that our subjects were doing much the same as Wickens et al.'s subjects.

The second-probe data are neither wholly consistent nor inconsistent with the Wickens et al. model. For the OO combination, the effects at the two-word set size are consistent with the view that the items are in PM on the second probe, in that the latency for the second probe on the delayed test was the same as the latency on the second probe on the immediate test. However, the pattern changes on the four-word set-size trials. Assuming that the items are in PM on the second probe, one would expect to find that latencies would be equivalent on the immediate and delayed tests, just as they were in the two-word set-size case. Clearly, those expectations have not been realized; latencies on the delayed trials were substantially longer than those on the immediate trials.

There are a number of possible explanations for this pattern of results. On the basis of the Halford et al.'s (1988) research, it is possible that, for our subjects, the capacity of PM is close enough to four items, so that with a four-item but not a two-item memory set, responding to the probe knocks one or more items out of PM. If this were to happen, a second retrieval would need to be carried out when the second probe of a four-item memory set was presented.

Our preferred explanation, however, is that the equivalent latencies for the two-word set size on the immediate and delayed tests reflect the use of a recall strategy. We think that after responding to the first probe, the other member of the set is not in consciousness but is brought back into consciousness by a deliberate recall strategy. The probe is then directly matched to the second item. This explanation seems reasonable given that on the NO

combinations, in which this strategy cannot be used, there was a significant effect of delay on second-probe reaction time for both the two-item and the four-item memory sets. Our data would suggest that the use of this strategy is limited only to the case in which two items are in the memory set, so that after an old probe, only a single item has to be brought back into consciousness as opposed to the multiple items in Wickens et al.'s proposal.

The difference in RT between the first and second probes is also problematic for the notion that the second-probe items are in PM. One possible explanation for the quicker reaction time to the second probe could be that, on some occasions, the list is in consciousness when the second probe is presented even though it had not been when the first probe had been presented. If this were the case, one would expect that the difference between the first and second probes would be greater on delayed tests than on immediate tests. However, the obtained data are the opposite to this prediction. The second-probe advantage was more pronounced on the immediate test than on the delayed test.

Whatever the explanation for the results, clearly there is only mixed support for the Wickens et al. model. We decided to test the model further by examining the effects of PI on the second probe.

EXPERIMENT 2

In Experiment 2, we looked at PI effects on the first and second probe, using delayed tests only. Again, we expected to replicate Wickens et al.'s (1981) data on the first probe by finding equivalent scanning rates in all conditions as well as intercept differences between the high- and low-interference trials. In addition, we predicted that intercept differences would disappear on the second probe if the list items are brought back into consciousness when the first probe is presented.

The procedure used in this experiment was conceptually the same as that used by Wickens et al. (1981), however some differences were introduced. Wickens et al. varied the number of items in the three trials per taxonomic category by using two formats: the first and third trials with set sizes of two and the second trial with a set size of four, or the reverse (i.e., using the first and third trials with set sizes of four and the second trial with a set size of two). In Experiment 2, we changed the Wickens et al. procedure by presenting all three trials per category with the same memory-set size. The Wickens et al. format was adapted to control the amount of PI on the third trial. That is, irrespective of the set size on the high-interference trial, six instances from the category had been seen prior to it. Whether or not this is the proper way to control for the amount of PI is problematic. We think that PI could be a function of the number of prior items, the number of prior lists, or an average of the number of prior items weighted by how recently the interfering material had been presented (see Humphreys & Tehan, 1988). Because we did not believe that the Wickens et al.

procedure necessarily controlled for the amount of PI, and because we were primarily concerned with whether the pattern of results was the same for both the first and the second probes, we decided to present all three trials per category with the same set size. This procedure may not equalize PI effects, but it does have a major advantage in that the data from the second trial in each group can now be incorporated into the analyses.

Method

Subjects

Two different cohorts of 20 psychology students each participated in this experiment. The first cohort were introductory psychology students who participated for course credit. The second cohort were third-year undergraduate psychology students who participated in the experiment as a class assignment. All 40 subjects were tested individually in a session that lasted approximately 1 h.

Materials

The stimulus pool used in this experiment was derived by sampling 18 instances from each of 36 different taxonomic categories. These categories and their respective instances were selected from Casey and Heath's (1988) Australian category norms, Nelson and McEvoy's (1982) norms, or Shapiro and Palermo's (1970) norms. In selecting the categories, care was taken to minimize between-category similarity.

The 18 instances from each category formed the basis of three trials. Of the six words randomly assigned to each trial, four were randomly assigned to be memory-set items (two of these were randomly selected on two-word set-size trials) and two were selected to be negative probes. Thus, in all, there were 108 experimental trials, comprising three consecutive trials from each category. The first trial in each set of three using the same category represented the low-interference condition. The second trial theoretically should increase the amount of interference, while the third trial in each set represented the high-interference condition.

The 108 trials were presented in blocks of 54 trials. One block of 54 trials involved memory-set sizes of two; the other experimental block consisted of 54 trials each with a memory set of four items. Thus, memory-set size was blocked rather than randomized as it had been in the first experiment. The first cohort of subjects received the two-word set-size trials first, followed by the four-word set-size trials. The second cohort of subjects received the four-word set-size trials first and the two-word set-size trials second.

Again, the same probe combinations used in the first experiment (Old-Old, Old-New, New-Old, New-New) were used here. Old probes were ones that were studied as part of the memory set. New words were words from the same category as those in the memory set, but had not been studied on the current or any previous trial. For the 18 low-interference trials in each block, 8 trials had the OO combination, 4 trials had the ON combination, 4 trials had the NO combination, and 2 trials had the NN combination. The same ratio of probe combinations was used in the second and third (or high-interference) trials using the same category. The actual probe combinations across the three trials using the same category were randomized. As in Experiment 1, positive probes were randomly sampled without replacement from the memory set.

Prior to the experimental blocks being presented, the subjects studied eight practice lists of unrelated words. Half the practice trials used memory sets of two and half used set sizes of four. In addition, each cue combination was used once for each set size.

Procedure

The procedure used in this experiment was almost identical to the delayed condition used in the last experiment. After a brief

“ready” signal, the subjects were presented with either two or four items to study. The first set of subjects read these words aloud; however, due to an oversight, the second cohort did not read the words aloud but studied them silently. Then a random three-digit number was displayed and the subjects counted backward in threes from this number for 12 sec. Then the screen went blank for 2 sec, after which the first probe was presented. After the subject responded to the first probe, the screen went blank for a further 2 sec, which was followed immediately by the second probe. After the subject responded to the second probe, an intertrial interval of 4 sec ensued.

Results

Because there were some procedural differences between cohorts, initial analyses were done in which cohort was a factor. The results of these analyses showed that the first cohort was reliably faster than the second on all probe combinations. We are not sure whether this difference is due to the second cohort's taking more care before making a decision, or to the fact that the first cohort read the items out loud whereas the second cohort read them silently. In addition, on the first probes, the set-size effect was much stronger for the second than for the first cohort. By definition, these set-size effects lead to changes in the slope of the scanning function. This difference in set-size effects appears to be due to asymmetrical practice effects associated with the different orders in which the conditions were presented. The first cohort was tested on the four-item trials after having practiced the procedure on 54, two-item trials. In contrast, the second cohort started the session with the four-item trials and was thus without the benefit of practice on the two-item trials. We believe that differential set-size effects reflect the differential practice each cohort received on the four-item trials. In any case, the critical feature of the data is that set-size effects were found in both cohorts. In addition, interference effects were essentially the same in both groups. To view the results without these practice effects, the data from both groups were collapsed and then analyzed. Again, median RTs were calculated only on trials in which the subjects had made the correct response. All analyses were 2×3 repeated-measures ANOVAs, with set size and interference condition as the factors.

First-Probe Analyses

Positive probes. An inspection of the results showed that the pattern of results found on the first probe approximates Wickens et al.'s (1981) data. Latencies for set sizes of two were faster than latencies for set sizes of four. The scanning rates across the three trials were roughly equivalent, and intercept differences emerged as PI built up. The only surprising finding was that there was almost as much PI on the second trial as there was on the third. Since this finding was common to all analyses, the data from Trials 2 and 3 have been collapsed in the figures.

Analysis of the latencies to correct responses to old first probes (OO and ON) revealed reliable effects for set size [$F(1,39) = 38.21$, $MS_e = 45,248$] and interference condition [$F(2,78) = 4.72$, $MS_e = 9,468$]. The interaction

between memory-set size and interference condition was not reliable.

Negative probes. The latencies from 1 subject were eliminated from the negative first probes (NO and NN) because that subject provided fewer than two correct responses for one condition. The latency data for the other 39 subjects were analyzed. The only reliable effect was that of set size [$F(1,38) = 34.04$, $MS_e = 45,498$]. The apparent absence of an interference effect is problematic in that we did not replicate Wickens et al.'s (1981) findings. As is evident in Figure 3, the pattern of results seems to be similar for positive and negative probes, but the effects appear to be stronger for the positive probes.

The accuracy data for both positive and negative probes produced only one reliable effect—that of memory-set size [$F(1,39) = 19.04$, $MS_e = 1.508$, and $F(1,39) = 9.02$, $MS_e = 0.312$, for positive and negative probes, respectively]. On the positive trials, the subjects were 94% accurate on trials with a set size of two and 89% accurate on trials with a set size of four. For negative probes, accuracy was 96% and 92% for set sizes of two and four, respectively. It would seem that speed-accuracy tradeoffs are not present, in that the subjects were slower to respond and made more errors with the larger set size than they did with the smaller one.

Second-Probe Analyses

Positive probes. The second probes of the OO and NO combinations were analyzed separately, in accord with Experiment 1. For the OO combination, the effect of memory-set size was reliable [$F(1,39) = 210.27$, $MS_e = 13,732$], as was the effect of interference condition [$F(2,78) = 5.05$, $MS_e = 8,491$]. The interaction between memory-set size and interference condition was not significant.

Although latencies to the second probe were considerably faster than latencies to the first probe, the overall pattern of performance was quite similar to the first-probe results, as can be seen in the top panel of Figure 4. Responses were faster when the set size was two rather than four, and the scanning rate was roughly equivalent in both conditions. The different intercepts indicate that PI had an effect on performance. The point is that we did not replicate the two-word set-size effects found in Experiment 1.

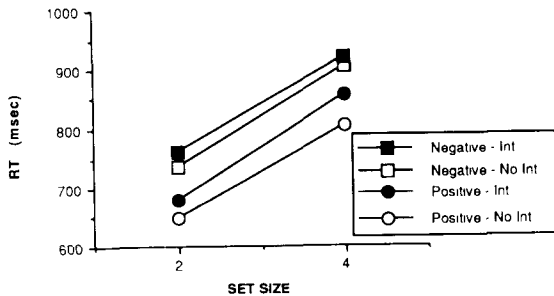


Figure 3. First-probe data for Experiment 2. Reaction time as a function of memory-set size, interference level, and nature of the probe.

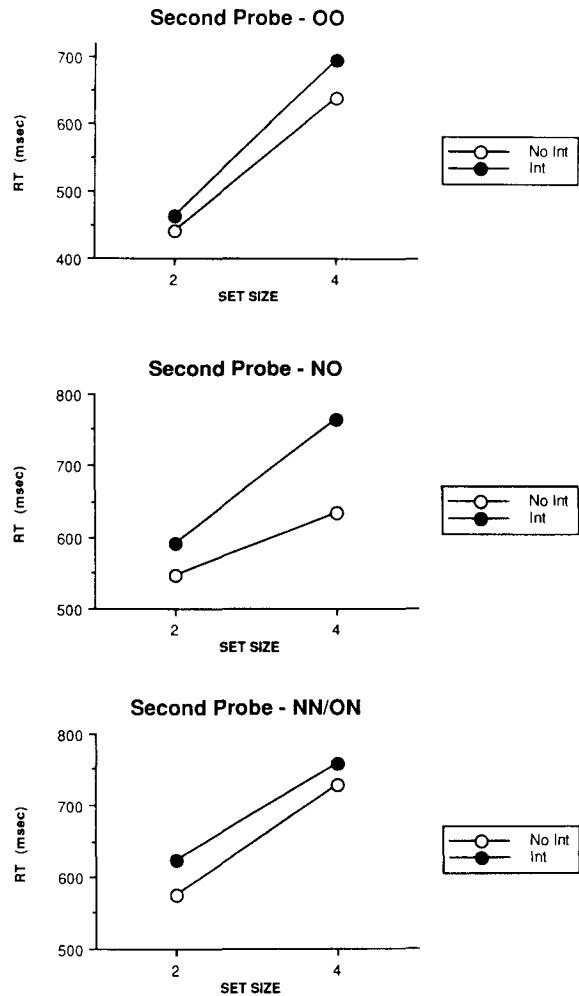


Figure 4. Second-probe data for the OO probe combination (top), the NO probe combination (middle), and the combined ON/NN probe combinations (bottom) in Experiment 2. Reaction time as a function of memory-set size, interference level, and nature of the probe.

In the analysis of the NO condition, 3 subjects' data were eliminated because they produced fewer than two correct responses in at least one of the conditions. Analysis of the latency data of the remaining 37 subjects revealed reliable effects of set size and interference [$F(1,36) = 19.60$, $MS_e = 56,528$, and $F(2,72) = 6.48$, $MS_e = 43,166$, respectively]. The interaction was not significant.

The middle panel of Figure 4 again shows that the basic pattern of results was maintained with the NO combination. Set-size effects are present, scanning rates are within tolerable limits, and PI effects are present.

Negative probes. As with the first probes, the data from the second probes from the ON and NN combinations were collapsed to get a reasonably reliable measure of performance. Again, the data from 1 subject had to be eliminated. Reliable effects for set size and interference condition were found [$F(1,38) = 81.08$, $MS_e = 14,168$, and $F(2,76) = 8.29$, $MS_e = 5,100$, respectively]. As

with the negative first probes, the effects of set size were present with the second probe. However, unlike the negative first probe, the effects of interference were readily apparent on the second negative probe, as can be seen in the bottom panel of Figure 4.

The accuracy data for the second probes in the OO, NO, and combined ON/NN probe combinations were analyzed. On OO trials, the subjects were 96% accurate on trials with a set size of two and 87% accurate on trials with a set size of four. For NO trials, accuracy was 94% and 88% for set sizes of two and four, respectively. Likewise, on negative trials, accuracy was 97% and 94% for set sizes of two and four, respectively. In all cases, the only reliable effect obtained was that of set size [$F(1,39) = 19.82$, $MS_e = 1.43$, $F(1,39) = 8.82$, $MS_e = .425$, and $F(1,39) = 6.39$, $MS_e = 0.319$, for OO, NO, and negative probes, respectively]. Again, speed-accuracy tradeoffs do not seem to be evident.

The subjects responded to the second probes more quickly than they did to the first probes. For the two-word set-size trials, the difference between first- and second-probe RTs was 173 msec on the first trial in blocks in which an old probe was being tested. The difference was 185 msec on the second trial and 181 msec on the third trial. On trials in which a new probe was presented, the second-probe advantage was 157 msec on the first trial in a block of three, 147 msec on the second trial, and 130 msec on the third trial. For the four-word set-size data in which an old probe had been presented, the second-probe advantage was 181 msec on the first trial, 187 msec on the second trial, and 198 msec on the third trial. For four-word set-size trials in which a new probe was presented, the difference was 190 msec on the first trial, 183 msec on the second trial, and 180 msec on the third trial.

To establish the reliability of the second-probe advantage, $2 \times 3 \times 2$ repeated-measures ANOVAs, with probe number, trial number, and probe type being the factors, were carried out on the two-word and four-word set-size data. For both set sizes, the second probe produced reliably faster latencies [$F(1,39) = 70.70$, $MS_e = 44,574$ for two-word set size, $F(1,39) = 35.19$, $MS_e = 115,689$ for four-word set size]. Again, the effects of PI were found for both set sizes [$F(1,39) = 9.53$, $MS_e = 5,650$ for the two-word set size, $F(1,39) = 4.82$, $MS_e = 14,189$ for the four-word set size]. Likewise, the subjects responded to old probes more quickly than they did to new probes [$F(1,39) = 79.24$, $MS_e = 13,175$ for the two-word set size, $F(1,39) = 22.32$, $MS_e = 27,326$ for the four-word set size].

Discussion

As with Experiment 1, the first-probe data replicate Wickens et al.'s (1981) data and are thus consistent with their model. Set-size effects were strong and the effects of PI were present and seem robust enough for us to conclude that the changes in procedure used had little effect upon performance. One somewhat surprising result found with our procedure, which was not testable using the

Wickens et al. procedure, is that PI was as strong on the second trial of each set as on the third. It would seem that PI builds up very rapidly (Humphreys & Tehan, 1988).

In Experiment 2, unlike Experiment 1, there was no support for the idea that list items remained active in PM prior to the second probe. For both the OO and the NO combinations, PI effects were present for both set sizes, suggesting that some sort of retrieval was taking place after the presentation of the second probe.

In Experiment 1, retrieval effects had not been present for the second probe in set sizes of two, but this was due predominantly to the OO combination. For this reason, we argued that, with set sizes of two, the subjects may have been using the strategy of recalling the item that had not been probed the first time around and then doing a direct match with the second probe when it was presented. It seemed plausible that the same strategy could be used in the current experiment and, if it were used, one could reasonably expect to see little effect of PI for set sizes of two items in the OO combination. In Experiment 2, however, the amount of PI on set sizes of two was approximately the same for the OO and NO combinations. It is possible that the subjects in Experiment 2 did not use the strategy or did not use it as often (the use of such a strategy in Experiment 1 may have been stimulated by the immediate tests on the two-word set size), or, alternatively, the finding in Experiment 1 may not be reliable. In any event, there is clearly no evidence for the hypothesis that retaining four items in PM while responding to the first probe strains the capacity of PM whereas retaining two items and responding does not. If this were occurring, we should not have found PI on the second probe for set sizes of two in Experiment 2.

As in Experiment 1, the subjects responded to second probes much more quickly than they had to first probes. The second-probe advantage seems to be fairly standard in this experiment and does not seem to be influenced in any way by the buildup of PI. At this time, our preferred explanation for the second-probe advantage is a general preparation to respond. During the 2-sec unfilled retention interval on the immediate tests in Experiment 1, the subjects were actively engaged in learning the list (rehearsing it or organizing it). Thus, they may have had insufficient time to prepare to encode and respond to the probe, or their ability to predict exactly when the probe would occur may have been impaired. On the delayed tests, there was also a 2-sec delay which was ostensibly unfilled. However, this interval was at least partially filled by the conclusion of the counting activity. In contrast, the 2-sec interval prior to the second probe started with the conclusion of the subject's response (the button press). Thus, there may have been more time to prepare or a greater certainty about the time at which the second probe would be presented than there was with the first probe. The increased preparation would account for both the faster responses and the fact that the results were somewhat more consistent with the second probe than they were with the first probe (the greater consistency being

due to a reduction in the variance). This explanation is supported by the failure to find any variable that influenced the magnitude of the second-probe advantage in Experiment 2.

It is, of course, possible that a portion of the second-probe advantage could reflect memory processes in addition to an enhanced preparation to encode and respond. To establish this would, however, require a control condition in which the first probe probes some other memory set (e.g., a fixed, long-term memory set). Support for a memorial explanation would be obtained if the response to the second probe was faster when the first probe probed the same memory set than when it probed a different memory set. Such a finding would have important implications for our understanding of probe recognition. The important point with respect to this paper, however, is that the second-probe advantage is not due to an enhanced presence in primary memory. If this were the case, the second-probe advantage in Experiment 1 should have been greater for the delayed condition than for the immediate condition. Instead, it was significant in the opposite direction.

GENERAL DISCUSSION

The first-probe data in both experiments are consistent with an explanation in terms of retrieving a list from SM and making it active in PM, followed by an exhaustive serial search through that list (Wickens et al., 1981, 1985). However, the second-probe results in both experiments are contrary to what one would predict assuming a reasonable extension of the Wickens et al. model. In Experiment 1, latencies were generally longer on the second probe in delayed trials than in immediate trials, thus indicating some act of retrieval from SM. In addition, in Experiment 2, in which only delayed tests were used but PI varied, reliable effects of PI were found in all second-probe conditions. Overall, the data suggest that if a variable affects the first probe (immediate vs. delay, low PI vs. high PI), it affects the second probe as well. This seems to be incompatible with Wickens et al.'s assumption that when the first probe is presented, the list items or a pointer to the list is brought back into consciousness.

The above conclusion is based upon the premise that the second probe is functionally an immediate test. One possible way of maintaining the viability of the Wickens et al. model is to deny the premise on which we have based our arguments. If the assumption is made that a second retrieval takes place between the first and second probes, the data are perfectly consistent with the Wickens et al. model. We do not think that such an assumption is warranted, however. It seems highly unlikely that a small set of items that were in an active state would no longer be in an active state just because a simple response had been made to a probe, especially when subjects knew that they would be probed again on these items in 2 sec time. Note in particular that Baddeley and Hitch (1974)

have shown that subjects could recall a list of three or four digits in order after having taken 2 or 3 sec to perform a sentence-verification task. A supporter of the second retrieval assumption should be prepared to explain why Baddeley and Hitch's subjects could maintain three or four digits in PM while doing a sentence-verification task, whereas our subjects could not maintain two words during the 1 sec it took them to respond to the probe.

If the second retrieval assumption is rejected, our data set can be added to a number of data sets that make it increasingly difficult to maintain the retrieval-plus-scan model for delayed probe recognition proposed by Wickens et al., or, for that matter, to maintain the scan model for immediate probe recognition proposed by Sternberg (1969). For example, both Monsell (1978) and Baddeley and Ecob (1973) have provided data that are more consistent with a direct access model for memory retrieval than they are with the serial scan proposed by Sternberg and adopted by Wickens et al. Monsell used the varied-set version of the Sternberg paradigm, varying it only in presentation rate (two items per second) and retention interval (the probe was presented immediately after the last item rather than after a 2-sec unfilled retention interval). The intent of these modifications was to control the subjects' rehearsal strategies. Monsell was interested in the serial-position effects of the positive probes as well as the effects of different types of negative probes. Discussion of the negative-probe results will be taken up later. For the positive probes, Monsell found that latency was a function of how recently the probe item had been studied. Moreover, he found that the serial-position curves could be superimposed. The latency for the last item was the same irrespective of set size, as was the latency for the second from the last item, the third from last, and so forth. This superimposition was only qualified by the faster RT to the first item in any list. He concluded that RT to positive probes could be described as a function only of recency (trace strength), minus a constant if the probe happened to come from the first serial position.

Baddeley and Ecob (1973) also employed the varied-set version of the Sternberg paradigm. Their modification of the task involved repeating items in the memory set. Thus, in the repeated-items condition of their study, subjects studied lists in which one or more items were repeated (e.g., 2474). These lists then were probed with positive or negative probes. They found that RT was faster to positive probes from lists that contained repeated items than from lists that contained no repeated items. They argued that this result was consistent with a trace-strength hypothesis, but could be consistent with the scanning notion if it was assumed that repeating an item had no functional effect on set size. That is, a list of four items with one repeated had a functional set size of three, whereas a list of four items with no repeats had both a nominal and a functional set size of four. If this were the case, one would expect a difference in latency for the negative probes in each case, since the functional set size differed.

Baddeley and Ecob found no such difference and hence rejected the scanning hypothesis in favor of a trace-strength hypothesis.

A further set of findings that seems hard to explain in terms of the retrieval-plus-scan model is the effect on RT when words from prior trials are used as probes on the current trial. In a number of studies (Halford et al., 1988; Monsell, 1978), the uniqueness of the probes has been varied across various trials. Halford et al.'s data are probably the most appropriate because they based their methodology on the Wickens et al. model. In addition, their negative-probe data replicate Monsell's negative-probe data in all respects. In the Halford et al. experiments, two types of probes were used on the high-interference trial of a set of three trials using ending categories. Their "new" probe was one that had the same ending as the words in the memory set but had not been seen on either of the first two trials from that particular category. Their "old" probe was a word that had been seen previously on one of the first two prior trials from that particular category. In each case, the probe could be positive or negative (i.e., it could have been present or not present in the current trial). Halford et al. found an interaction between new and old probes on positive and negative trials. On the negative trials, the latency for old probes was reliably slower than that for new probes, and accuracy was lower on old probes than on new probes. This was true of all memory-set sizes. However, when the old probe was a positive probe, subjects were more accurate than they were for new probes but were faster to respond to old than to new probes. Unlike the negative trials, however, this pattern was not maintained for all set sizes (G. S. Halford, personal communication, 1988). One of the explanations for these results, suggested by Halford et al., is that subjects occasionally search previous sets. This leads to more appropriate responses on positive than on negative trials. Accepting this explanation preserves the scanning notion, but it has the ramification that if subjects are searching previous sets as well as the current one, the slopes and intercepts become uninterpretable. That is, the values for the slopes and intercepts will depend on the probability that a previous list is searched (this could vary with positive and negative probes) and the time needed to search the previous list (this will vary with the length of the previous lists). The importance of this observation is that it destroys the entire evidentiary base for assuming a serial scan in the first place.

In our opinion, the second-probe data from the current set of experiments show that the evidence implicating a retrieval operation for the entire list is extremely weak. Without such a retrieval operation, the idea that on delayed probe-recognition tests memory access is gained via the kinds of scanning operation envisaged by Wickens et al. (1981) must be discarded. Instead, it appears that memory access must start with the probe item itself and that the same memory-access operation must occur on the first

and second probes. Furthermore, these results join with those of Baddeley and Ecob (1973), Monsell (1978), and Halford et al. (1988) in casting doubt on memory scanning as an explanation for the immediate-test results. The problem here is that the same set-size effects are being obtained on an immediate test, when items are plausibly available to be scanned, and on a delayed test in which, without the postulated retrieval operation, the items are clearly unavailable for scanning. It thus appears that a memory-access operation that starts at the presentation of the probe is required for both immediate and delayed tests.

The postulation of a memory-access operation that starts with the probe provides a convenient explanation as to why nothing happened during the 2-sec unfilled interval between the end of counting and the presentation of the probe on the delayed tests. This is a problem for the retrieval-plus-scan model (see the discussion in Wickens et al., 1981) because it seems plausible that subjects would use this unfilled interval to retrieve the list so that it would be in consciousness when the probe was presented. The problem with adopting the assumption that memory access starts with the probe and discarding the retrieval-plus-scan model is that doing so also entails discarding a very convenient explanation of why the effects of PI are so similar for probe recognition and recall (Wickens et al., 1981). A possible solution to this problem comes from the global matching models reviewed by Humphreys, Pike, Bain, and Tehan (1989).

These models differentiate between a matching operation used for recognition and a retrieval operation used in recall. In the matching operation, the matches between the probe and all of the list or experimental memories are combined into a single value. Thus, these models can account for the kind of repetition effects found by Baddeley and Ecob (1973), Halford et al. (1988), and Monsell (1978). Furthermore, the matching operation can involve only a single cue or a pair of cues. In all of these models, when a pair of cues (e.g., *AB*) is matched against memory, the result is similar to retrieval using *A* as a cue. Our proposal, then, is that probe recognition is a matching operation and that pair matching is applied to categorized lists. That is, with categorized lists, it is assumed that a composite cue consisting of the category label of the last list and the probe item is matched against memory. At this time, these models do not predict latencies, and they do not incorporate the kinds of assumptions needed for the recency effects found in these paradigms. Thus, it is not possible to use them to make precise predictions about probe-recognition paradigms nor is it possible to use them to explain why PI effects disappear at the shortest of retention intervals. Nevertheless, the distinction between matching and retrieval is sufficiently well defined in these models to be testable. Furthermore, the proposal that probe recognition with categorized lists is a form of pair recognition and is thus similar to cued recall has the potential to explain both the similarities and the differences between recognition and recall results.

REFERENCES

- BADDELEY, A. D., & ECOB, J. R. (1973). Reaction time and short-term memory: Implications of repetition effects for the high-speed exhaustive scan hypothesis. *Quarterly Journal of Experimental Psychology*, **25**, 229-240.
- BADDELEY, A. D., & HITCH, G. (1974). Working memory. *Psychology of Learning & Motivation*, **8**, 47-89.
- CASEY, P. J., & HEATH, R. A. (1988). Category norms for Australians. *Australian Journal of Psychology*, **40**, 323-339.
- CRAIK, F. I. M., & JACOBY, L. L. (1975). A process view of short-term retention. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (Vol. 1, pp. 173-192). Hillsdale, NJ: Erlbaum.
- CRAIK, F. I. M., & LEVY, B. A. (1976). The concept of primary memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 4, pp. 133-175). Hillsdale, NJ: Erlbaum.
- CROWDER, R. G. (1982). The demise of short-term memory. *Acta Psychologica*, **50**, 291-323.
- FRIENDLY, M., FRANKLIN, P. E., HOFFMAN, D., & RUBIN, D. C. (1982). The Toronto word pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods & Instrumentation*, **14**, 375-399.
- HALFORD, G. S., MAYBERY, M. T., & BAIN, J. D. (1988). Set-size effects in primary memory: An age-related capacity limitation? *Memory & Cognition*, **16**, 480-487.
- HUMPHREYS, M. S., PIKE, R., BAIN, J. D., & TEHAN, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix and TODAM models. *Journal of Mathematical Psychology*, **33**, 36-67.
- HUMPHREYS, M. S., & TEHAN, G. (1988). *A simultaneous examination of recency and cueing effects*. Unpublished manuscript, University of Queensland, St. Lucia, Queensland, Australia.
- JAMES, W. (1890). *Principles in psychology* (Vol. 1). New York: Holt.
- MONSELL, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, **10**, 465-501.
- NELSON, D. L., & MCEVOY, C. L. (1982). Category name and instance norms for 106 categories of various sizes. *American Journal of Psychology*, **95**, 581-634.
- PAIVIO, A. V., YUILLE, J. C., & MADIGAN, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology: Monographs*, **76**(1, Pt. 2).
- SHAPIRO, S. I., & PALERMO, D. S. (1970). Conceptual organization and class membership: Normative data for representatives of 100 categories. *Psychological Monograph Supplements*, **3**(11, Whole No. 43).
- STERNBERG, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, **57**, 421-457.
- WATKINS, O. C., & WATKINS, M. J. (1975). Build-up of proactive inhibition as a cue overload effect. *Journal of Experimental Psychology: Human Memory & Learning*, **104**, 442-452.
- WICKENS, D. D., MOODY, M. J., & DOW, R. (1981). The nature of timing of the retrieval process and of interference effects. *Journal of Experimental Psychology: General*, **110**, 1-20.
- WICKENS, D. D., MOODY, M. J., & VIDULICH, M. (1985). Retrieval times as a function of memory set size, type of probes, and interference in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 154-164.

(Manuscript received August 17, 1988;
revision accepted for publication April 5, 1989.)