A Bottleneck in Memory Retrieval from a Single Cue

Timothy C. Rickard Harold Pashler
University of California, San Diego

Abstract

We explored whether two memory retrievals from a single cue can be completed in parallel. Subjects were trained to make a vocal response and a key press response for each of a set of ten visually presented words. Subsequently, these two tasks were interleaved with a third, dual task condition in which subjects retrieved both the vocal and the key press response on each trial. Findings include: 1) the latency of the first retrieval on dual retrieval trials exceeded that of either of the two tasks performed by themselves, 2) the time to finish the dual retrieval task was roughly the sum of the time to finish the two tasks on separate trials, 3) the correlation between time to finish the first task (RT1) and the inter-response interval (IRI; latency between the first and the second response) was not significant, and 4) some subjects always choose to execute the slower task first on dual task trials. Taken together, these and other findings favor a sequential retrieval model over a broad class of parallel models.

One goal of cognition research is discovery of principles for predicting when two or more processes can or cannot be executed in parallel. This paper examines the case of memory retrieval. More specifically, the paper explores whether two independent memory retrievals from a single cue can be completed, from stimulus perception to response execution, in parallel. This question is fundamental to the study of memory and attention. Yet, remarkably, it appears that no study has ever directly addressed it. Our goals are to document the basic empirical phenomena, and to test candidate theories.

Several studies in the literature may provide insight into this issue. Rohrer, Pashler & Etchegaray (1998) investigated category exemplar retrieval in the case in which all items to be retrieved are members of a single category, compared to the case in which subjects must alternate retrievals between two categories.

In Experiment 1 of their study, subjects studied four exemplars from each of two categories (e.g., for the category animal, they studied the words goat, horse, pig, and sheep). On some trials, subjects were instructed to recall the examplars from only one of the categories. On other trials, they were instructed to alternately recall exemplars from each of the two categories. The IRIs were much slower on trials on which recall alternated between categories than on trials on which recall occurred only within a single category. Based on these and related results, the authors argued that items from different categories cannot be retrieved in parallel. Their conclusion regarding within-category recall, however, was quite different. IRIs progressively increased from the first to the last exemplar recalled within a given trial, and this effect was greater in the alternating category than in the single category condition. The authors showed mathematically that this effect is can be explained if one assumes that within-category retrieval is parallel, though they did not conclusively eliminate other possible accounts.

The between-category condition in Rohrer et al. (1998) can be seen as analogous to a dual task design, since two separate cues (one corresponding to each category label) were presented at the outset of each trial. There is converging evidence from dual task studies employing the psychological refractory period paradigm (Welford, 1952) that parallel retrieval is not possible when there are two separate cues on each trial (Carrier & Pashler, 1995; Anderson and Lebiere, 1998). In contrast, the within-category trials of Rohrer et al. can be seen as somewhat analogous to the design of our current experiment, in which only a single cue (e.g., a single category label) is presented, and two or more retrievals are performed using that cue. Their results suggest that parallel retrieval may be observed in the current experiments. Note, however, that being of the same category, retrievals in the within category condition were not independent in the sense that they are in the current study

The results of a study by Ross and Anderson (1981) provide additional evidence

Send correspondence to: Timothy Rickard Department of Psychology, 0109 9500 Gilman Drive University of California, San Diego San Diego, Ca 92093 – 0109 trickard@ucsd.edu

suggesting that two memory retrievals from a single cue can be completed in parallel. They had subjects study lists of paired associates. Across items in the list, each cue word appeared several times, in each case paired with one of two response words (e.g., baker-garage, or baker-canyon). For all cue words, one of the response pairings occurred frequently in the list (the strongly associated response), and the other occurred infrequently (the weakly associated response). In a subsequent test phase, subjects saw the cue words presented one at a time, with instructions to recall the first response word that came to mind. Not surprisingly, the strong associates were retrieved more often and more quickly than the weak associates. Of more interest, the upper tails of the response time (RT) distributions for strongly and weakly associated response words appeared to converge, a phenomenon identified previously by Townsend (1974) as consistent with parallel retrieval. In their task, however, subjects were only required to complete one retrieval per trial. Their results thus do not bear directly on the question of whether two retrievals can run to completion and be executed in parallel. Their results suggest only that, in their task, both responses were activated in parallel prior to execution of the first response.

Although both the Rohrer et al. (1998) and the Ross and Anderson (1981) papers suggest that retrieval of two responses from a single cue might be completed in parallel, recent studies by Rickard (1997; 1999) suggest the opposite. Rickard showed that the strategy shift from use of multi-step algorithms to direct memory retrieval in tasks like mental arithmetic can be naturally accounted for assuming that only one of these strategies can be executed to completion on a given performance trial (cf. Logan, 1988; Nosofsky & Palmeri, 1997; Palmeri, 1997; 1999). The model that he proposed assumes that only one response can be retrieved per attempt and also that most multi-step algorithms used in skill learning tasks reflect one or more long-term memory retrieval steps. It follows from that model that two strategies cannot be executed in parallel. It also follows that two responses cannot be retrieved in parallel given a single cue, or given two different cues for that matter.

Thus, there is a theoretical tension in the literature regarding the case of retrieval from a single cue. The findings of Rohrer et al. (1998) and Ross and Anderson (1981) indicate that retrieval of two responses might be completed in parallel. The Rickard (1997; 1999) studies suggest the opposite. This paper is intended to help resolve this tension.

In experiment 1, subjects were first trained to make a vocal response (saying a single digit) and a key press response for each of a set of ten visually presented words. The ten word stimuli were mapped to ten vocal (digit) responses, and to two (left or right) key press responses. Subjects were then required to perform a dual task in which they made both the vocal and the key press response on each trial (single-task blocks were interleaved with the dual-task blocks in this phase of the experiment).

Model Predictions

Two contrasting simplest case models will be considered first: an unlimited capacity independent parallel retrieval model (i.e., a race model), and a sequential retrieval model. As we note later, there is reason to believe that both of these models are too simple and will need elaboration. Nevertheless, we test the simplest case models first both to facilitate communication of the basic principles, and to confirm that augmented versions of these models are needed. The race model assumes that, in the dual task condition, the two retrievals can be initiated and completed independently and in parallel. According to this model, both retrievals can be executed concurrently at their normal speed and accuracy. The sequential retrieval model, on the other hand, assumes a bottleneck in central retrieval processes (i.e., processes in between perceptual and motor processing), such that one retrieval operation must wait while the other is executed. This model is consistent with the proposal of Rickard (1997), and can also be seen as a generalization of the central processing bottleneck model (for a review, see Pashler, 1994) to long-term memory retrieval (see also Carrier sand Pashler, 1995). Both of the models assume that perception, memory retrieval, and motor components of processing are stochastically independent serial processes, p, r, and m, respectively.

The Race Model

According the race model, there is first a single perceptual event, p, for the stimulus, followed by independent parallel retrieval and motor execution for the two tasks. Provided that subjects execute each response as soon as it becomes available, this model predicts that:

$$\mu_{RT1} = \mu_p + \mu_{[\min(RTkr + RTkm\ ,\ RTvr + RTvm)]} \quad (1)$$

$$\mu_{RT2} = \mu_p + \mu_{[\max(RTkr + RTkm\ ,\ RTvr + RTvm)]} \quad (2).$$

Here, μ refers to the population mean of a response time distribution, RT1 is the latency to the first response made on a dual task trial (regardless of which task is completed first), RT2 is the latency to the second response made on a dual task trial, (RTkr + RTkm) is the combined retrieval and motor component latency of a key press trial, and (RTvr + RTvm) is the combined retrieval and motor component latency of a vocal trial. The terms min(RTkr + RTkm , RTvr + RTvm) and max(RTkr + RTkm , RTvr + RTvm) refer to the prediction that, if one key press and one vocal response are drawn randomly from their respective distributions, the minimum, or smaller, of the pair constitutes a theoretical observation for RT1 for the corresponding dual task item, and the maximum (or larger) of the pair constitutes a theoretical observation for RT2.

Estimates for the expected value of RT1 and RT2 on each dual task trial can be generated from the single task data by determining the minimum and maximum of the corresponding single task vocal and key press RTs on the immediately preceding single task blocks (we assume that single task learning during the test is negligible). Consider, for example, the case in which a subject is presented with the cue “Red.” A random sample from the theoretical distribution for RT1 under the parallel model can be obtained by simply selecting the faster of the key press and vocal single RTs for that cue from the immediately preceding single task test block for each task. Analogously, an estimate for RT2 can be obtained by selected the slower RT. By averaging these minimum and maximum RTs over items and subjects just as for RT1 and RT2 in the dual task case, estimates of the parallel predictions for μ_{RT1} and $\mu_{RT2,}$ respectively, can be obtained. These estimates are slightly biased, as compared Equations 1 and 2, in that RTp is implicitly included as part of the parallel processing component rather than as a processing component common to both tasks that occurs prior to the onset of parallel processing. Stated as equations, the estimates generated from the single task data correspond to:

$$\mu'_{RT1} = \mu_{[\min (RTp + RTkr + RTkm\ ,\ RTp + RTvr + RTvm)]} \quad (3)$$

$$\mu'_{RT2} = \mu_{[\max (RTp + RTkv + RTkrm,\ RTp + RTvr + RTvm)]} \quad (4)$$

Equations 3 and 4 will tend to yield estimates for μ_{RT1} and μ_{RT2} that are slightly smaller than should be the case according to Equations 1 and 2. However, simulations reported in Nino and Rickard (in press) indicate that the difference in predictions of Equations 1 and 2 versus Equations 3 and 4 should be no more than about 10 ms, and thus can be ignored in the following analyses without compromising theoretical inference.

In addition to the point estimates given above, the race model predicts the following two inequalities:

$$\mu_{RT1} < \mu_{RTk}, \quad (5)$$

$$\mu_{RT2} > \mu_{RTv} \quad (6)$$

where RT_v is the overall latency (i.e., includes perceptual, retrieval, and motor components) for a vocal response on a single task trial and RT_k is the overall latency for a key press response on a single task trial. To simplify exposition, we assume here that the key press task is, on average, the faster of the two single tasks. This assumption will be verified by the data for each subject in Experiment 1.

The predictions in Equations 5 and 6 follow directly from the essential properties of race models (for discussions, see Colonius & Ellermeier, 1997; Colonius & Vorberg, 1995; Compton and Logan, 1991; Diederich & Colonius, 1987; Logan 1988, 1992; Miller, 1982; Rohrer, Pashler, and Etchegaray, 1998; Schweikert, 1983; Townsend & Ashby, 1983; Townsend & Colonius, 1997; Townsend & Nazawa, 1995), and can be understood by considering two cases for RT_1. On most dual task trials, the key press task (with the shorter RT distribution) will finish first. On these trials, RT_1s are distributed just as they would be if the key press task were executed by itself. On the remaining trials, however, the vocal response will be executed faster than the key press response. That is, on some trials, the key press RT will happen to be unusually slow, whereas the vocal RT will happen to be unusually fast, leading to reversal of the expected task RT ordering, with RT_1 corresponding to the vocal task. In this second case, RT_1 will always be somewhat shorter than would have been the case if the key press were executed by itself. Thus, the upper tail of the RT_1 distribution is compressed when the two tasks are performed on the same trial, relative to when the key press task is performed by itself, leading to a faster mean for RT_1 than for RT_k. This result is expected provided only that the RT distributions for the two tasks overlap. The reasoning for RT2 is analogous to that for RT1. On some trials, the key press RT will happen to be long, and the vocal RT will happen to be short, leading to a reversal of the expected task RT ordering. These trials result in a compression in the lower tail of the RT2 distribution (pushing it up), relative to the RT distribution for the vocal task performed by itself. It follows that the mean of RT2 will be longer than the mean of RTv.

The Sequential Model

Provided that subjects execute each response as soon it becomes available, and assuming for the moment that task scheduling (i.e., determination of which task to complete first) requires zero latency, then the prediction for the mean RT1 made by the simplest case sequential model is:

$$\mu_{RT1} = \mu_p + (j)(\mu_{kr} + \mu_{km}) + (1 - j)(\mu_{vr} + \mu_{vm}) \quad (7)$$

where j is the proportion of trials on which the key press task is completed first in the dual task condition. Thus, RT1 is a weighted average of RTv and RTk, with a weighting factor, j. An estimate of μ_{RT1} for a given dual task trial was obtained by noting which task was completed first on that dual task trial, and selecting the RT of the corresponding single task item (key press and vocal) on the immediately preceding vocal and key press single task blocks. Consider, for example, a dual task trial on which the key press task is completed first. At the trial level, j takes a value of either 1 (key press first) or 0 (vocal task first). Thus, the RT1 prediction for that trial reduces to: $\mu_{RT1} = \mu_p + \mu_{kr} + \mu_{km}$, which is directly estimated by the observed single task key press RT on the preceding single task block. The subject task ordering for each dual task trial was used to set the value of j (0 or 1) for the prediction based on the immediately preceding single task trials. Thus, j is not a free parameter in the model fits; the sequential model, like the race model, has the virtue of being parameter free. Note that because μ_{RT1} averaged over items and subjects is a weighted average of the vocal and key press latencies, its value must be between μ_{RTk} and $\mu_{RTv.}$

There is evidence in the literature that central processing, such as memory retrieval, can take place in parallel with, and without interference from, motor processing of another task (Pashler, 1997). Incorporating this evidence as an assumption in the sequential model (which predicts that processing must be sequential only within the central, retrieval stage of processing), the prediction for RT2 is:

$$\mu_{RT2} = \mu_p + \mu_{kr} + \mu_{vr} + \mu_{mv} \quad (8)$$

An estimate of this prediction for a given dual task trial was obtained by summing the corresponding single task RTs on the immediately preceding key press and vocal single task blocks. Again under the assumption of stochastically independent additive processing, the average of these summed RT corresponds to:

$$\mu'_{RT2} = (\mu_p + \mu_{kr} + \mu_{km}) + (\mu_p + \mu_{vr} + \mu_{vm}) \quad (9)$$

This equation differs from Equation 8 in it has one extra μ_p component (i.e., perception is counted twice) and that it includes the motor component for the first completed task (whichever that may be on a given trial), which according to the sequential model need not contribute to the observed reaction time since it can take place in parallel with the retrieval stage of the second completed task. Thus, the sum of the single task RTs is biased to over estimate μ_{RT2} in two respects. If we make a reasonable assumption that cue perception requires around 100 ms, and both key press and vocal motor output also require about 100 ms, then Equation 9 overestimates μ_{RT2} by about 200 ms (note **Hal's quarterly report finding**). Despite this potential for relatively substantial bias in the μ_{RT2} prediction, for simplicity and to make a fair comparison to the simplest case race model, we used the unadjusted sum of the single task RTs, corresponding to Equation 9, in the model fits to RT2 for Experiment 1. We will return this issue later.

General Assumptions for Both Models

The predictions of both models rest on the assumption that the single tasks are represented and processed independently, in both functional and stochastic senses, on the dual task trials. Conditions for meeting the functional independence assumption are almost certainly present on the first dual task trial for each item, since the component single tasks were learned and performed independently and in different practice phases. Independence may well hold over all five dual task test blocks in the experiment below, but that possibility cannot be guaranteed a priori. The special case of component task independence is theoretically important. Most potentially applicable parallel models in the literature (e.g., Logan, 1988; Palmeri,1997; Wenger, 1999) assume that parallel retrieval can take place in this case, although functional independence is certainly not a necessary condition for all types of parallel models.

An implicit assumption of our approach to fitting the dual task data is that subjects use the same stopping rule on single and dual task blocks, and that subjects are treating the tasks the same in all other respects on dual and single task blocks. It is possible that this assumption is false. For example, subjects could adopt a stricter stopping rule for the dual task, requiring more information accrual prior to executing responses than for single task trials. Different stopping rules could distort dual task RTs, potentially compromising theoretical inference. Fortunately, this possibility can be tested. If subjects did change their stopping rule in the manner outlined above, it would likely be observable as a speed-accuracy tradeoff such that accuracy is higher for the dual task compared to the single tasks.

Finally, note that the RT1 and RT2 predictions of both models can be viewed as both point predictions and lower bounds. The strongest prediction of both models is that RTs cannot be below the predictions (the lower bounds). If RTs are significantly below one or more of these boundaries, then the corresponding model(s) can be rejected. Note that the sequential model should be easier to falsify on these grounds, since its boundaries are substantially higher than those of the parallel model. A somewhat weaker prediction of both models is that dual task RTs should match the point estimates. This prediction is weaker because there is no a priori reason to expect that subjects will perform whatever process is available to them at their maximum possible efficiency on all trials, and particularly on the first few trials. That said, the closer the dual task RTs are to the model predictions, the stronger the support for the model.

Experiment 1

Method

Subjects

Sixteen University of California at San Diego undergraduate students participated for course credit.

Material, Design, and Procedure

Subjects were tested on IBM-compatible personal computers running Micro Experimental Laboratory software (version 2.01) and a button box and voice key (model 200A) all manufactured by Psychology Software Tools (Pittsburgh, PA). Stimulus materials are shown in Appendix A. All stimuli were presented in the standard MEL white font, at the center of a 14" monitor screen. Eight subjects were tested using each stimulus-response set.

Subjects were tested individually. At the outset of the experimental session, each subject was seated about 50 cm from the monitor. The program for phase 1 of the experiment was then initiated and the experimenter read aloud the instructions presented on the screen while the subject read along silently. The instructions informed subjects that their task was to memorize a unique single digit number that the experimenter had randomly paired with each color name, and that they would later be required to recall and speak the correct digit into a microphone when presented with a color word.

The first two blocks of stimulus presentation were always "study" blocks (one block consisted of one presentation trial for each of the ten color stimuli, randomly ordered). On each trial of each the study block, a single color name (e.g., "red"), the answer (e.g., "4"), and instructions to memorize the answer, all appeared on the screen simultaneously, with the answer centered 2 lines below the stimulus, and the instructions centered two lines below the answer. After 5 seconds, the instruction to memorize the answer was replaced by instructions to make the correct response when ready. After the subject responded, the computer proceeded with presentation of the next item.

After the second study block was completed, the subjects were asked whether they felt sufficiently comfortable with the task to proceed to the trials on which they would be required to produce the answers themselves. If they responded "no," they were allowed only one additional study block. Subjects then performed a series of production blocks, in which they were again shown the color names one at a time at the center of the screen, but this time without the answers. They were required to speak the answer they had memorized into the microphone as quickly as possible while being as accurate as possible. After the subject responded and the voice key tripped, the experimenter entered the digit corresponding to each subject's response using the numerical key pad and then pressed the "+" key. If the voice key failed to trip as the subject vocalized their response, the experimenter entered the subject's response, but then pressed the "-" key to provide a record of trials on which voice key failures occurred. The computer then provided accuracy feedback and, if the subject was in error, presented the correct response. These blocks continued until the subject performed 2 consecutive blocks with 100% accuracy and with an average RT on the last block of less than 1200 ms.

In Phase 2, subjects memorized either a left- or right-side button response for each of the same ten color names seen in Phase 1. This phase proceeded as in Phase 1, with the exception that the microphone was moved away from the subject, who was instructed instead to respond by using the pointer fingers of the left and right hand to press the far left or far right buttons (respectively) on the button box.

In Phase 3, the test phase, subjects received interleaved blocks requiring vocal, key press, or both vocal and key press responses. The vocal and key press blocks in this phase were identical to those of the previous phases. On blocks requiring a vocal response, subjects placed their elbows near the edge of the table and folded their hands over their arms (this was the posture in phase 1 as well). This procedure kept the subject a roughly constant distance from the microphone and monitor. On trials requiring only a key press response, the microphone was moved outside of the field of view.

If we denote the single task blocks as the S condition, the dual task blocks as the D condition, the vocal task with v, and the key press task with k, then the sequence of blocks in phase three was: S_v, S_k, S_v, S_k, D, S_v, S_k, D, S_v, S_k, D, S_v, S_k, D, S_v, S_k, D, S_v, S_k. Note that there were four S blocks (2 of each task) before the first D block. The first two of these S blocks was included as a warm-up to re-familiarize subjects with the single tasks.

Brief instructions were presented on the computer screen and read by the experimenter at the beginning of each phase 3 block. Instructions for the single task blocks during phase 3 simply informed subjects that they were doing the same tasks as in the first two phases, and reminded them to respond as quickly as possible while maintaining high accuracy. Prior to each dual task block, subjects were again instructed to perform both tasks as quickly as possible while maintaining high accuracy. These neutral instructions should not result in any bias toward the use a particular strategy, other than promoting use of the fastest and most efficient strategy for performing the tasks. On correct trials in phase 3, subjects received no feedback. On incorrect trials, they received feedback and were informed of the correct answer(s), just as in the first two phases.

Eight of the subjects received exactly the sequence of tasks outlined above. The remaining 8 subjects received the reverse order, such that the key press task was given in phase 1, the vocal task in phase 2, and the ordering of key press and vocal tasks were reversed in phase 3. Four subjects within each of these two groups of 8 learned stimulus-response mapping A, and the other 4 subjects in each group learned stimulus-response mapping B (see Appendix A).

Results and Discussion

Phases 1 and 2

Four trials (0.68%) on which the voice key failed to correctly trip were excluded from the following analyses. Subjects required on average of 5.43 generate blocks to reach the learning criteria for the vocal task, and 3.18 generate blocks to reach the learning criteria for the key press task. The minimum number of generate blocks required was 2 for both the vocal and key press tasks. The maximum was 12 and 11 for the vocal and key press tasks, respectively. The grand mean RT on the last generate block was 970 ms for the vocal task, and 690 ms for the key press task.

Phase 3: Test

The eight trials (0.45 %) on which the voice key failed to trip correctly were excluded from subsequent analyses. The first vocal and key press single task blocks, which were included as warm-up blocks, and were also not analyzed. Thus, the data analyzed below consist of 6 single task blocks for each task, and, interleaved between these, 5 dual task blocks.

Overall accuracy was high. Averaging across all test blocks, the proportion correct for the vocal task performed by itself was .976, and for the key-press task performed itself was .970. Accuracy was nearly identical on the dual task trials: Accuracy for the vocal task was .968, and for the key press was .966. The proportion of dual task trials on which both tasks were performed correctly was .935. Both theories assume single task independence on dual task trials. Thus, both theories predict that the proportion of dual task trials on which both responses are correct will be the product of the accuracies of the single task trials. The product of the single task accuracies was .928, quite close to the corresponding dual task accuracy. It does not appear that subjects adjusted their speed-accuracy criteria in the dual task phase, and thus there is no reason to suspect that subjects used different stopping rules in the single and dual task blocks.

All RT results were computed for correct trials only. Figure 1 shows the overall means on each block for RT_k, RT_v, RT1, and RT2 (means were computed first over items within subject and then averaged over subjects), along with predictions of the race and sequential models for RT2. Several patterns are clear. First, consistent with the training results, key press task was performed faster on average than the vocal task, and this finding held for each of the 16 subjects. Second, performance in both of the single task conditions remained stable over the practice blocks. A within subjects ANOVA revealed no significant changes in mean RT as a function of block for either the vocal task, $F(5, 75) = 1.73$, $p = .14$, or the key press task, $F(5, 75) = .93$, $p = .47$.

Third, with the exception of the first dual task block, RT_2 was close to the sequential model prediction derived from the sum of the single task RTs. An ANOVA with factors of Block and Type (observed RT2 versus predicted RT2), revealed no main effect of Type, $F(1, 15) = .56$, $p = .46$, and this finding held with the first test block removed, $F(1, 15) = .13$, $p = .72$. In contrast, RT_2 was significantly slower than predicted by the parallel model, $F(1, 15) = 6.78$, $p = .02$, and this finding also held with the first test block removed, $F(1, 15) = 11.45$, $p = .004$.[1]

The fourth pattern in Figure 1 is that RT1 was substantially slower than RTv. This result violates the predictions of both the parallel model (Equation 5) and sequential model (Equation 7). One possible account of these unexpectedly slow RT1 values is that some subjects adopted a strategy of grouping the responses in the dual task conditions into a single “output event,” rather than outputting each response as soon as it became available. For such subjects, RT1 would include not only the time to retrieve the first response, but also any additional time required to retrieve (or finish retrieving) the second response. Conversely, their IRI’s would be extremely short, limited only by the degree of precision with which the two responses can be locked into a single output event. Although this grouping strategy, which was observed in previous research (Borger, 1963; Pashler & Johnson, 1989), is not predicted by either model, it is in principle consistent with both. That is, regardless of whether retrieval itself is sequential or parallel, some subjects may simply find it preferable to synchronize the two responses into a single action. Clearly Equations 1 and 7 do not describe the predictions of either model for subjects who group their responses. Thus, to test the model predictions for RT1, it was necessary to identify those subjects likely to have adopted that strategy and to remove them from the analysis. This was the primary goal of the individual subject RT analyses described next.

Individual Subject RT Results

Mean RTs for the single and dual tasks and for the sequential and parallel predictions for RT2 were computed for each subject and task, collapsed across test blocks, as shown in Table 1 (along with other subject-level data). The mean for RT2 was above the sum of the single tasks for nine of the 16 subjects. Two subjects, numbers 4 and 12, had a mean RT2 more than 200 ms below the sum of the single tasks. However, the mean RT2 for all subjects was near or above the sum of the single tasks on the first test block (values not shown in Table 1). Important individual differences are apparent in the IRI’s. First, subjects 2, 3, 4 and 8 all had extremely short mean IRI’s of less than 100 ms, despite RT2s that were consistent with the prediction of the sequential model. These IRIs strongly suggest that these subjects adopted a grouping strategy in the dual task condition. A second cluster of five subjects (1, 10, 12, 13, and 15) exhibited IRI’s that were between 100 and 300 ms. Although these IRI’s are probably too long to justify a pure grouping interpretation, grouping may well have occurred on at least some trials for these subjects. A final cluster of seven subjects (5, 6, 7, 9, 11, 14, and 16) exhibited mean IRI’s that were in all cases longer than 344 ms. The grand mean IRI collapsed over these 7 subjects was 554 ms. Inspection of each test block for these subjects revealed none on which the mean IRI was faster than 300 ms. Thus, it seems reasonable to conclude that these subjects, who we will refer to as non-groupers, executed each response in the dual task condition as soon as the answer became available, making no effort to group responses. Nino and Rickard (in press) found even stronger evidence for bimodality in the IRI’s over subjects in a closely related experiment. Thus, it is difficult to dismiss this effect as anything other than individual differences in grouping strategy. Even if one prefers a parallel retrieval account, **it is quite unlikely that xx subjects over four** experiments (see also, Rickard and Nino, in press) would have mean IRI’s of less than 100 ms, many of which were less than 50 ms. Individual differences in response grouping strategy seems an unavoidable conclusion, regardless of one’s broader theoretical stance.

The test results limited to the seven non-groupers are shown in Figure 2, along with predictions of the sequential model for RT1 and RT2. To enhance readability, predictions for the parallel model are not shown. That model predicts that RT2 is about 30 ms greater than RTv, and that RT1 is slightly less than RTk. Clearly both of these predictions are incorrect for the non-groupers, just as in the overall data set. The sequential model again under-predicts RT2 on the first couple of test blocks, but provides good match for the remainder of the test. However, RT1, while much closer to the sequential prediction for these subjects, was still above it throughout the test, suggesting that the simplest case sequential model cannot fully account for dual retrieval performance.

A Modified Sequential Model

The finding that the mean of RT1 for non-groupers was slower than expected by the sequential model is comprehensible, and probably to be expected, given a nearly ubiquitous phenomenon known in the dual task literature as the preparation effect (see Pashler, 1997, for a review). In nearly all dual tasks exhibiting a central processing bottleneck, the latency to make the first response in a dual task is slower, by at least 100 ms, than that to make the same response for that task performed by itself. Similar slowing is found on blocks on which the subjects perform only one of the tasks, but do not know which it will be (Pashler, 1998). Subjects in our task clearly must prepare to make two responses. Thus, it is reasonable that we should find the same preparation effect that is found in traditional dual task paradigms. Given this fact, a more realistic sequential model prediction for RT1 is:

$\mu_{RT1} = \mu_p + \mu_{prep} + (j)(\mu_{kr} + \mu_{km}) + (1 - j)(\mu_{vr} + \mu_{vm})$ (10)

where μ_{prep} might be approximately 100 ms (see **Pashler, xxxx, p. yyy**). Support for this modified account is garnered from the fact that the RT1 delay for non-groupers on the last test block, relative to the value predicted by the modified sequential model in Equation 10, is only about 50 ms.

On the earlier test blocks, the RT1 delay is much more substantial, suggesting, from the standpoint of the sequential model, that subjects were performing the task inefficiently on early test blocks. One factor that may well result in inefficient performance on initial dual tasks blocks is task scheduling; that is, selection of which task to complete first. We have assumed for simplicity that task scheduling has zero latency. Scheduling may well take time, however, and may be a learned skill, whose latency reduces with practice. As task scheduling latency decreases over test blocks, RT1 should approach the prediction of Equation 10, as was observed. A unique prediction of this inefficient scheduling account is that although the scheduling effect must occur prior to RT1, its effects on reaction time must propagate with the same magnitude to RT2. Hence, the under prediction of the sequential model for RT2 cannot be less than that for RT1. The data for non-groupers in Figure 2 do show this pattern. The majority of the under-prediction of the sequential model for RT2 can be understood as a consequence what is happening prior to RT1.

As noted in the introduction, two other modifications to the sequential prediction for RT2 are motivated by task analysis and by other empirical results. First, once subjects have encoded the single stimulus for the first retrieval, there is no reason to believe that they have to repeat this perceptual encoding before executing the second retrieval, as is assumed by fitting the RT2 data based on the sum of the single task RTs (Equation 9). If we assume that perception requires at least 100 ms **(brief justification)**, then this value needs to be subtracted from the prediction generated from the single task data. This 100 ms correction would roughly offset the effect of the dual task preparation delay for RT2 (but not RT1), leaving us again with Equation 8 as an approximate prediction for RT2. However, another factor that would be expected to yield savings in RT2 is a temporal overlap in motor processing for the first task with central processing (memory retrieval) for the second task (see Pashler, 1997, for discussion of central and motor processing overlap). If our estimate of 100 ms savings corresponding to the task 1 motor processing stage is subtracted from the RT2 prediction based on the single task data, then the estimate (based on the single task data) of the RT2 prediction for fully modified (i.e., including the 100 ms estimates for both the perceptual and motor corrections) sequential model is:

$\mu`_{RT2} = \mu_{prep} + (\mu_p + \mu_{kr} + \mu_{km}) + (\mu_p + \mu_{vr} + \mu_{vm}) - 200$ ms (11).

If the perceptual and motor adjustments are correct, then Equation 11 can be reduced to the theoretical modified sequential prediction for RT2,

$\mu_{RT2} = \mu_p + \mu_{prep} + \mu_{kr} + \mu_{vr} + \mu_{mv}$ (12).

Note that Equation 12 is simply Equation 8 with the μ_{prep} component added. Note also that the μ_{RT2} estimate derived through Equation 11 is only 100 ms faster than the RT2 estimate used in Experiment 1, which was obtained by simply added to two single task RT's, with no adjustment. Thus, the net effects of the modifications to the sequential model are simply to add 100 ms to the μ_{RT1} prediction, and to subtract 100 ms from the μ_{RT2} prediction.

The modified sequential prediction is about 40 ms below the observed mean RT2 values on the last three dual task test blocks shown in Figure 1. For non-groupers in Figure 2, the RT2 prediction is about 20 ms below the mean RT2 value on the last test block. When all plausible and straightforward adjustments to the sequential model are taken into account, the match of its predictions to the data is good, provided that one allows for inefficiencies in dual task processing on the first few dual task blocks.

Candidate Modifications to the Parallel Model

Adjustments can also be made to the race model in an effort to improve its fit to the data. First, the preparation cost, μ_{prep}, that was added to the sequential prediction is equally applicable to the parallel model, resulting in a 100 ms increase in both the RT1 and RT2 predictions. This modification, however, only modestly to improves the model's fit, given the roughly 400 ms difference between the parallel prediction and RT1 (averaged over the five test blocks), and the roughly 600 ms averaged difference between the parallel prediction and RT2.

Based on these, as well as other results to be discussed, the pure race model can safely be eliminated from consideration. This conclusion may not be surprising to the reader. Nevertheless, it is important with respect to the literature, because most parallel models to date assume a race. That assumption may be made primarily for expediency, facilitating derivation of analytical predictions while also keeping the number of free parameters low. That very fact, however, illustrates the value of rejecting the model. With the special case of a race eliminated, the parallel theorist must venture into the more complex set of models that incorporate a capacity limit or related mechanism to slow the retrieval process for both RT1 and RT2. For that class of models, computer simulations will likely be needed to demonstrate sufficiency in actual fits to the data. Further, it may prove difficult to build such a model that can fit the data well while also being sufficiently constrained to be empirically testable. The value of falsifying the race, then, is largely to motivate researchers to abandon that special case and instead explore the more general class of parallel models.

In the simplest type of limited capacity parallel model, capacity allocation to each task on each trial is independent of task or other factors related to retrieval (i.e, on average, each task gets 50% of the available capacity; there is no preference for one task over the other). This case is important to evaluate first, because it has the advantage over the sequential account of not requiring a task (or capacity) scheduling mechanism. It is also the least flexible form of a limited capacity model, though it does require at least one free parameter to find the optimal degree of capacity limit for fitting the data.

Because the key press task was performed much faster than the vocal task on average for all subjects, a limited capacity account (with no capacity scheduler) predicts that that task will usually be completed first on dual task trials (certainly on the majority of these trials for each subject). This prediction holds for ten of the subjects. However, subjects 5, 6, 9, 11 14, and 15 (five of these non-groupers) completed the vocal task first on 100%, 98%, 92%, 94%, 100%, and 81% of the trials, respectively. The race model simply cannot account the data from these subjects, nor can the limited capacity account, given the assumption that capacity allocation is independent of task. A limited capacity model can only account for performance of these six subjects by adding a task scheduling mechanism, such that the vocal task can strategically be given the majority of capacity, allowing it to finish before the key press task. We shall refer to this model as the <u>scheduled limited capacity</u> parallel model. Such a model would constitute a hybrid between sequential and limited capacity parallel models, because it would incorporate a capacity scheduling operation analogous to that of task selection component sequential model. In the sequential model, task scheduling is by definition a dichotomous process (either one task or the other). In a scheduled limited capacity model, capacity can be allocated to each task over a continuum of values. The scheduled limited capacity account thus has greater flexibility with respect to fitting the RT data. It requires at least two free parameters: one for setting the capacity limit and one for capacity allocation (i.e., capacity scheduling). Further, it appears that the scheduled limited capacity model can only fit the data for non-grouper subjects who completed the much slower, vocal task first on nearly all trials by allocated nearly all capacity first to the vocal task. In that special case it reduces to the near equivalent of the sequential model.

As noted previously by Rohrer, et al. (1998), a parallel retrieval model generally predicts a negative correlation between RT1 and the IRI. If the two retrievals occur in parallel, and if RT1 on a particular trial is unusually long, then on average the IRI will be unusually short (because RT2 performance would not be expected to be especially slow on that trial). The reverse is also true. Hence, a parallel model, whether capacity limited or not, predicts a negative correlation between RT1 and the IRI. Rohrer et al. (1998) showed through simulations that the expected magnitude of the correlation under in the case of a race is about -0.5. The sequential processing model also assumes that the two retrievals are independent, but crucially of course, it assumes that they are performed sequentially. Recall also that non-groupers nearly always selected one of the tasks (key press of vocal) first. Under these conditions, the sequential model predicts zero correlation between RT1 and the IRI.

To test the predictions above, we preformed a general linear model analysis on the dual task data separately for the seven non-grouper subjects (each of whom completed who either the key press of vocal task first on at least 92% of the trials), treating the IRI as the dependent variable, block and item as categorical covariates, and RT1 as a continuous variable of interest. The effect of RT1 on the IRI was not significant for any of the seven subjects (p-values ranged from .075 to .90). Further, the mean partial correlation due to IRI over the seven subjects was +.065, a result which is in the reverse direction of the parallel model prediction. Again, the results are clearly inconsistent with the race and limited capacity accounts. The scheduled limited capacity model might account for this lack of correlation between RT1 and IRI, but again only if the majority of capacity is allocated to one task at-a-time.

The limited capacity model without a scheduling component may be able to account for the performance of the five grouper subjects, but only if combined with the strategic grouping that we have assumed. Alternatively, it is possible that the five grouper subjects are not intentionally grouping outputs, but rather are, for some mysterious reason, allocating just enough extra capacity to the vocal task that the two responses are executed at the same item.

Neither the unscheduled nor scheduled capacity limited parallel retrieval account predicts a priori that mean RT2 value averaged over all subjects will be equivalent to or above the sequential prediction, as was observed. This result carries important information. Based on both the literature, and, we suspect, intuitions of researchers in the field, the most likely outcome space for the dual task RTs would be between the race prediction and the sequential prediction, with a bias toward the race prediction. A priori, it should have been a simple matter to falsify the sequential model if it were wrong.

Of course, the dual task RTs were above the lower bound predictions of the race model as well. One might argue that the same performance inefficiency argument that we applied to the sequential model can also be applied to the parallel models. Several factors, however, make an inefficiency account applied to those models far weaker. With respect of the race model, the mean RT fits were far worse than those of the sequential model. Performance inefficiencies would have to be several hundred milliseconds more severe in the race account. Even on the last test block, they would have to be on the order of 400 ms. Second, by the end of test, the sequential fits were quite good. We interpreted this pattern as reflecting the near elimination of dual task performance inefficiencies. This interpretation is supported in a more recent study by Nino and Rickard (in press; Experiment 2). They studied extended dual task practice in a very similar task. For non-grouper subjects, they found that the mean RT2 values converged near the sequential prediction after the fist few test blocks, and then traced that prediction for all remaining test blocks (there were 30 in total). The IRIs also converged and traced the sequential prediction throughout the test. These patterns held not just for the mean, but for the entire RT distribution. Clearly the RT lower bound predictions of the sequential model are meaningful. Third, whereas an initial performance inefficiency is theoretically motivated for the sequential model, due to the need to select which task to complete first, there is no reason to expect such inefficiencies in the race and limited capacity models, beyond the dual task preparation effect. By definition in those models, there is no task scheduling operation. If a capacity scheduler is added, there could be scheduling inefficiencies just as for the sequential model, but it would seem odd, in our view, for a capacity scheduling delay to yield RTs no faster than predicted by purely sequential retrieval. Finally, two independent results noted earlier, the findings that five subjects completed the slower task first on nearly all trials and that the RT1, IRI correlations were non-significant, speak against the parallel class of models on other grounds.

The sequential model is completely consistent with the correlation results, provides by far the best account among the models tested of the mean RT results, has zero degrees of freedom, and does not encounter difficulty in accounting for the five subjects who always completed the slower, vocal task first. In fairness, the sequential model as developed here really does not make any predictions regarding which task subjects will chose to execute first. With respect to this variable at least, the race and limited capacity models (but not the scheduled limited capacity model) are more constrained.

The reader may question the fairness of treating the sequential model as having zero free parameters, compared to at least two for the scheduling limited capacity model. After all, for the sequential model to fit the RT data exactly, even for non-grouper subjects, there would need to be at least one parameter to account for the much greater decrease in dual task than single take RTs over the course of test. Unfortunately, this criticism ignores the fact that one or more free parameters would also be needed for any of the parallel models we have considered to accommodate that result. If, for example, the single task RTs remain constant over test blocks, then the parallel model we have evaluated must, like the sequential model, predict that the dual task RTs will remain constant as well, a prediction which is clearly false. If we assume that the number of parameters required to accommodate the greater rate of speed up for the dual task would be the same for the sequential and parallel models, then it is correct to conclude that the sequential model needs at least two fewer free parameters than the most competitive, scheduled limited capacity model.

Experiment 2

Although the results of Experiment 1 favor a sequential over parallel retrieval accounts, replication across different tasks is needed before strong general conclusions can be reached. One potentially important feature of Experiment 1 is that the stimulus-response mappings for the two tasks were different. For the vocal task, each of the ten color names was mapped onto a unique digit response. In contrast, for the key press task, the mapping was from ten stimuli to only two responses. We presumed that making the key press task simpler, and thus its execution faster, would maximize our chances of observing parallel retrieval if it were possible. However, it is possible that that very design feature worked against parallel retrieval. Perhaps the performance bottleneck apparent Experiment 1 did not reflect an intrinsic, fundamental constraint on memory access. Instead, it could be that some type of interference is caused when subjects attempt to perform two retrievals simultaneously while also coordinating two different stimulus-response mappings. In this special case, it is conceivable that parallel processing is artificially prevented, or its behavioral expression somehow masked. If so, a symmetrical mapping should allow for parallel dual task performance. In the following experiment we tested this possibility by using eight color word stimuli, eight vocal-digit responses, and eight different key press responses. Each response corresponded to exactly one stimulus.

Method

Subjects

Nineteen University of California at San Diego undergraduate students participated for course credit.

Material, Design, and Procedure

Methods were identical to those of Experiment 1, with the following exceptions. First, there were eight color word stimuli, as opposed to ten stimuli used in Experiment 1. Second, each stimulus was associated with one of eight different key press responses. The keys corresponded to the standard keyboard keys for Q, E, F, V, M, K, O, and [. A red dot was affixed to each key to indicate which set of keys were the appropriate responses. Prior to initiation of each key press or dual task trial, subjects placed eight fingers, excluding thumbs, on the keys in the natural left to right order. The vocal response for each stimulus in this experiment was a consonant letter, with each stimulus requiring a different letter response.

Results and Discussion

Data from phases 1 and 2 were not analyzed. Each subject reached the single task learning criteria and was able to complete the experiment in less than one hour. In phase 3, the voice key failed to trip on 3.5 % of trials, and these were excluded from subsequent analyses. As in Experiment 1, the first vocal and key press single task blocks, which were included as warm-up blocks, and were not analyzed.

Overall accuracy was high. Averaging across all test blocks, the proportion correct for the vocal task performed by itself was .97, and for the key-press task performed itself was .965. Accuracy on the dual task was .979 for the first task completed, and .969 for the second task completed. The proportion of dual task trials on which both responses were correct was .948, and the corresponding prediction based on assumption of task independence was .936.

All RT results were computed for correct trials only. Figure 3 shows the means on each block for RT_k, RT_v, RT1, and RT2, along with RT2 predictions for the modified sequential and modified race models, and the RT1 prediction for the sequential model. Equating the stimulus-response mapping pattern for the two tasks brought vocal and key press single task RTs much closer together than in Experiment 1, the grand mean for the vocal task being 1171 ms and that for the key press task being 1080 ms. However, the performance on the key press task was still faster on average for 14 of the 19 subjects**.**

The dual task RTs, relative to model predictions, were analogous to those of Experiment 1. RT2 was initially above the modified sequential prediction, but by the fourth and fifth test block stabilized at a level approximately equivalent to it. RT2 was also well above the modified race prediction (Equation 2 with μ_{prep} added). RT1 was again above the modified sequential prediction of **xxx ms, and to a lesser extent above the modified race prediction of yyy ms.** In the ANOVA's on RT2, there were main effects of both Type, $F(1, 18) = 6.51$, $p = 0.02$, and the Type X Block interaction, $F(1, 18) = 5.78$, $p < 0.01$, for the comparison data to the modified sequential model. With the first block, removed, however, neither the effect of type, $F(1, 18) = 1.74$, $p > 0.20$, nor the Type x Block interaction, $F(1, 18) = 0.83$, $p < .0.20$, were significant.

The RT2 values were again significantly slower than predicted by the parallel model. For Type, $F(1, 18) = 56.32$, $p < 0.001$, and for the Type X Block interaction, $F(1, 18) = 5.05$, $p < 0.002$. These effect of Type was still significant even with the first test block removed, $F(1, 18) = 32.1$, $p > 0.20$. The effect of the Type by Block interaction, however, was not, $F(1, 18) = 1.13$, $p > 0.20$.

Overall, subjects made the vocal response first on 43% of the trials. As in Experiment 1, order of responding varied greatly over subjects, with 6 subjects making the key press response first more than 90% of the time, and 4 subjects making the somewhat slower vocal response first 90% or more of the time. For these later four subjects, and scheduling component is clearly required in any limited capacity model. Because the RT distributions of the two single tasks overlap much more than in Experiment 1, it is not immediately obvious whether the remaining subjects' response orderings can be accounted for under a limited capacity model without a capacity scheduler. However, the subset of the single task data that constitutes the parallel model prediction provides an estimate of the percentage of vocal task first trials expected for the dual task. That is, for each pair of single task trials corresponding to a subsequent dual task trial, the minimum RT provides not only the parallel RT prediction, but also information about which task should win the race. Thus, the proportion of data pairs from the single task on which the faster RT corresponded to the vocal task constitutes the prediction for the proportion of dual task trials on which the vocal task should be completed first. If dual task retrieval is parallel with limited capacity divided evenly between the two tasks, which is implied if there is no capacity scheduler, then the proportion of trials on which the vocal task is finished first should be the same as predicted by the race model. For sample data, about half should be above and about half below the predicted proportion. For 18 of the nineteen subjects, however, the proportion vocal first trials on the dual task was more extreme (i.e., closer to 0 or 1) than predicted by the parallel model. This result is quite rare, with a p-value of less than .001 using a binomial test. As for Experiment 1, it appears that a limited capacity model without a capacity scheduler cannot fit the data.

Subject level results are summarized in Table 2. There was again a mixture of response order strategies. Seven subjects had means IRIs that never fell below 549 ms on any dual task dual task block, qualifying them as non-groupers based on the criterion described in Experiment 1. Fourteen subjects could not be classified as either groupers or non-groupers for all five test blocks. Only one subjects had a mean IRI (89 ms) across all test blocks small enough to qualify as a response grouper. However, there was a clear trend among the remaining subjects to move toward a grouping strategy over the course of the five test blocks. By the last test block, seven subjects had a mean IRI of no more than 110 ms, with the average among these subjects being 73 ms. The only other mode in the IRI distribution for the last test block was for four subjects having mean IRI's between 500 and 700 ms. Apparently subjects gravitate toward a response grouping strategy as they gain more experience with the task (for discussion of this and other dual task practice effects, see Nino and Rickard, in press).

For 14 of the 19 subjects, the mean of RT2 across the five test blocks was higher than predicted by the modified sequential model. Looking only at the first test block, where task independence is nearly guaranteed to hold, 17 of the 19 subjects had mean RT2 values higher than the modified sequential prediction, and ten remaining two subjects had mean RT2 values no more than 108 ms below that prediction. Keeping in mind the decreased stability of subject level means, the sequential prediction is successful as a lower bound performance level for both group and subject level data.

The mean RT results for the seven non-grouper subjects are shown in Figure 4. The difference between RT1 and the sequential prediction for it by the end of test is smaller than for the overall data, but the reduction is not as dramatic as for Experiment 2. It may be that performance inefficiencies are eliminated at a slower rate for this more complex mapping involving a separate key press response for each stimulus, though we have no independent evidence to support this hypothesis. As predicted by the task scheduling inefficiency account discussed in Experiment 1, however, the under-prediction of the sequential model for RT1 is again less than that for RT2.

Subject level correlations between RT1 and the IRI, using a procedure identical to that described in Experiment 1, were performed on data from four subjects who did not group responses and who executed either the vocal or the key press response first on at least 95% of the trials. For three of these subjects there was no correlation (all p's > 0.5) and for the fourth there was a significant positive correlation ($p > 0.02$). Across the two experiments, there were ten non-significant correlations, a result uniquely consistent with the sequential model, and one positive correlation, a result not consistent with either the sequential or the parallel model.

General Discussion

We explored the straightforward, but previously undressed, question of how people manage two independent memory retrievals from a single cue. Previous research in related task domains suggested that subjects might be able to handle such a task with grace, retrieving and executing both responses in parallel with little interference, and with consequent fast RTs. Our findings are more consistent with the opposite case. Both the RT results and informal post-experiment reports indicate that the task was initially daunting. Subjects almost unanimously reported that the dual task was initially more difficult than they had expected. Such comments suggest to us that subjects attempted parallel retrieval at first but found it impossible. RT results showed that there was generally no savings in performing the two tasks on a single trial relative to performing them on two separate trials. That is, dual task RT2 values on the first couple of test blocks were no faster than the sum of the component single task RTs. Indeed, on the first test block, dual task RT2 value were notably slower than the summed single task RTs.

A race model can clearly be eliminated from consideration for both experiments, based on the mean RT results, the RT1 vs. IRI correlation results, and the fact that a number of subjects nearly always executed the slower of the two tasks first on dual task trials. Any limited capacity parallel model which predicts that dual task RTs will below the sequential prediction, or that does not include a task scheduling mechanism, is also eliminated from consideration. A limited capacity account with capacity scheduling (and at least two free parameters) is potentially consistent with the results, although data fitting with a formally specified version of that model would be needed to convincingly demonstrate sufficiency. It also appears that such a model can only fit the data from at least some subjects if all capacity is allocated to one task at a time. In this case that model reduces to the near equivalent of the sequential model.

In all cases, mean RT1 and RT2 values were above, or not statistically different from, the predictions of the modified sequential model. Further, with the exception of the first test block in both experiments, the RT2 data were not significantly different from the sequential prediction, though there was a clear trend for better convergence with it on later test blocks. As we stated earlier, there is no a priori reason to believe that subjects will perform the dual task at maximum efficiency the first time they are asked to do so. Indeed, uniquely for the sequential and scheduled limited capacity models, there is reason to expect that they will not. On the first few test blocks there may be some delay in deciding which task to complete first. With practice, subjects may adopt a preference for retrieving the response for one of the task first, either at the item or task level, potentially reducing the time required to schedule task execution. There has in fact been a trend in all experiments we have run on this topic for subjects to move toward executing a favored task first a greater percentage of the time on the last few test blocks.

<u>Integration With Other Results in the Literature</u>

Our results appear to be at odds with those of Rohrer et al. (1998) regarding within-category retrieval. They concluded that retrieval of exemplars within a given category -- a special case of retrieval from a single cue -- can operate in parallel. One account of this discrepancy is that there is a fundamental difference between retrieval of two exemplars from a pre-existing category and retrieval of two independently learned responses from a single cue. We will touch on this possibility again later. Alternatively, there are candidate sequential processing accounts of their within category results. More empirical work is needed to firmly establish the properties of within-category retrieval, and to relate these two rather different paradigms.

Ross and Anderson (1981) also concluded that retrieval of two responses from a single cue can occur in parallel. Their task, however, never required an actual response for more than one item. One approach to reconciling their results with ours would be to assume that, although two responses can be retrieved in parallel, two retrievals cannot concurrently run to completion. Retrieval might involve two stages: a race between competing responses, followed by suppression of all but the most activated response. A second retrieval would then be necessary to access the second response. This response bottleneck model is consistent with the parallel activation of candidate first task responses that Ross and Anderson inferred. It is also consistent with recent dual task study by Logan and Schulkind (2000) using a semantic categorization tasks. In those tasks, first task responding was facilitated when the second task required the same response, implying parallel flow of activation from both task stimuli to the first response.

This response bottleneck account, however, does not fit well to our data. This model's prediction for RT1 is the same as that of the modified race model (Equation 1, plus a dual task preparation allowance), a prediction which is off by about 300 ms. In addition, the savings due to the race effect for RT1 should propagate to RT2, so that the model's prediction also falls about 300 ms below the observed RT2. A version of that model in which the first response is characterized by a race can therefore be rejected. The model could be revived by assuming limited capacity parallel retrieval for RT1, allowing the RT1 prediction to be raised to the observed values, while raising the RT2 prediction by the same amount, bringing the fit roughly in line with the data. However, the model faces the same problems already noted for the limited capacity model. Only a limited capacity account with a capacity scheduler is potentially consistent with the mean RT results, though it faces problems on other grounds.

Turning the tables, the sequential model does not appear to be consistent with the Ross and Anderson (1981) or Logan and Schulkind (2000) results, which suggest parallel retrieval up to at least the first response. One potentially important difference between our tasks and Ross and Anderson's is that in our tasks the two responses for each stimulus were learned under two different tasks sets, one being something like "press the appropriate key," and the other being "make the correct vocal response." In the Ross and Anderson experiment, both responses to each word were contained within the same study list, and they had the same output modality. Thus, it may be more appropriate for their case to think of learning of the two responses for each cue as having taken place under the same task set. With respect to task set, then, the two responses for each cue in our experiments were independent in a way that theirs were not. Task independence in this sense may prove crucial in determining whether or not retrieval is parallel, either for the first or second response. Consistent with this proposal, Logan and Schulkind (2000; Experiment 2) found no cross-talk (i.e., no difference in performance for trials with compatible vs. incompatible responses) in their dual task experiments only for the case in which the two tasks had different tasks sets.

Roher et al. (1998) within category results might be explained by a somewhat different violation of task independence. Since in that condition all responses on each trail were from a common category, task independence in the sense assured in our experiments may not have been present. Logan and Schulkind's first task facilitation effect (as measured by the effect of stimulus onset asynchronicity in the compatible condition) also occurred only when both retrievals were from the same category. Further exploration of the nature and performance consequences of task independence is important, but beyond the scope of this paper. For a more formal and integrative candidate model, see Rickard and Bajic (in press) and Rickard and Bajic (2002).

Our results also have implications for recent debates regarding the nature of dual-task interference. Whereas central bottleneck models beginning with Welford (1952) have generally assumed that interference reflects structural (although not necessarily immutable) processing limitations, some writers (e.g., Meyer, Kieras, Lauber, Schumacher, and others, 1995) have recently proposed that the interference might instead be strategic. In most dual-task experiments, they point out, subjects are strongly encouraged to complete the first response as fast as possible. While this might not logically demand postponement of the second task, they argue that subjects might nonetheless adopt a highly conservative strategy of completing the central processing in the first task before turning to the second task to be certain that responses do not come out in the "undesirable" order. The experiments described in this paper did not involve any instructions that could reasonably be claimed to favor one particular order of responding over the other. Further, in a more recent unpublished experiment, we replicated Experiment 1, manipulated instructions between subjects. In one of the instruction conditions, subjects were told to retrieve both responses at the same time. They were told that previous results show that if they concentrate, they should be able to do this. The results were nearly identical to those of Experiment 1. Finally, the use of a single stimulus as a cue for both responses makes it impossible to attribute the interference to peripheral factors of the sort that Meyer et al (1995) have invoked to explain other dual-task results. Therefore, the very strong central interference observed here provides a new challenge to these views, converging with other recent dual task studies that have dispensed with the traditional emphasis on speed of first-task responding and still found evidence for queuing of central processing (e.g., Ruthruff & Pashler, submitted; Carrier & Pashler, 1995).

Conclusions

Prior to these experiments, virtually nothing was known directly about the psychology of retrieving two independent responses from a single cue. Our results identify a number of clear, and a priori non-obvious, performance phenomena. These include: 1) RT2 (time to finish both retrievals) was equal to or greater than the sum of the RTs for the two tasks when performed by themselves. Given the near consensus in the field in favor of parallel retrieval models, and the numerous examples of facilitation in other, non-retrieval task domains, this result must surely be counted as surprising. 2) There were major individual difference in first task RT performance, such that some subjects executed each response as soon as possible, whereas others did not. 3)

Response ordering was stereotyped in both experiments. In many cases, subjects preferred to give the slower, vocal response first. At the least, this finding poses a significant challenge to parallel models in which there is no strategic allocation of capacity. 4) There was no evidence of any negative correlation between RT1 (time to finish the first retrieval) and the inter-response interval (IRI) for non-grouper subjects in either experiment. Again, parallel models have difficulty accounting for this finding. 5) The findings above were not idiosyncratic to a particular stimulus-response mapping scheme.Over the course of the paper, we set up and tested several plausible models. Some, including the simplest forms of the parallel and serial models, and straightforward limited capacity parallel retrieval models, can now be eliminated from consideration. In the end, we narrowed the pool to two possibilities: a modified version of the sequential model, and a limited capacity parallel model that incorporates capacity scheduling. Our rationally-based preference is for the former. The sequential model accounts for most of the data quite well, with the main exception being the mean RT results toward to beginning of test. We advanced a reasonable and theoretically motivated account of this finding that proposes that task selection requires time, but can become more efficient with practice. Since the sequential model has no free parameters, its inability to fit to the mean RT data perfectly in this first effort should not be surprising. The model has no freedom to contort its fit to mask inconsistent data patterns. This is a positive feature, in our view, since it gives us deeper insights into the merits and limitations of its underlying theoretical assumptions.

Author note

This research was supported by National Institutes of Health grant number 1 R29 MH58202-01A1 to the first author.

References

Anderson, J. R. & Lebiere, C. (1998). Cognitive Arithmetic. In Anderson, J. R. (Ed.), The Atomic Components of Thought. Lawrence Erlbaum: New Jersey.

Borger, R. (1963). The refractory period and serial choice-reactions. Quarterly Journal of Experimental Psychology, 15, 1 - 12.

Carrier, L. M. & Pashler, H. (1995). Attentional limits and retrieval from long-term memory, Journal of Experimental Psychology: General, 113, 518 – 540.

Colonius, H. and Ellermeier, W. (1997). Distribution inequalities for parallel models of reaction time with application to auditory profile analysis, Journal of Mathematical Psychology, 41, 19-27.

Colonius, H. and Vorberg, D. (1994). Distribution inequalities for parallel models with unlimited capacity, Journal of Mathematical Psychology. 38, 35-58.

Compton, B. J., & Logan, G. D. (1991). The transition from algorithm to retrieval in memory-based theories of automaticity. Memory & Cognition, 19, 151 - 158.

Diederich, A. and Colonius, H. (1987). Intersensory facilitation in the motor component. Psychological research, 1987, 23-29.

Hommel, B. (1998). Automatic stimulus-response translation in dual-task performance. Journal of Experimental Psychology: Learning, Memory , and Cognition, 24, 1368-1384.

Logan, G. D. (1988). Toward an instance theory of automatization. Psychological Review, 95, 492 - 527.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 883 - 914.

Logan, G. D. & Schulkind, M. D. (2000). Parallel memory retrieval in dual task situations: I. Semantic Memory. Journal of Experimental Psychology: Human Perception and Performance, 26, 1072-1090.

Meyer, D. E. & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: I. Basic mechanisms. Psychological Review, 1997, 104, 3-65.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. Cognitive Psychology, 14, 247-279.

Nino, R. and Rickard, T. C. (2001). Practice effects on two retrievals from a single cue. Manuscript submitted for publication.

Nosofsky, R.M., & Palmeri, T.J. (1997). An exemplar-based random walk model of speeded classification. Psychological Review, 104, 266-300

Palmeri, T., J. (1997). Exemplar similarity and the development of automaticity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23, 324-354.

Palmeri, T. J. (1999). Theories of automaticity and the power law of practice. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25, 543-551.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. Psychological Bulletin, 116, 220 – 244.

Pashler, H. (1997). The Psychology of Attention. Cambridge, MA: MIT Press.

Pashler, H. & Johnston, J. C. (1989). Interference between temporally overlapping tasks: Chronometric evidence for central postponement with or without response grouping. Quarterly Journal of Experimental Psychology, 41A, 19 – 45.

Rickard, T. C. (1999). A CMPL alternative account of practice effects in numerosity judgment tasks. Journal of Experimental Psychology, Learning, Memory, and Cognition, 25, 532-542

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. Journal of Experimental Psychology: General, 126, 288 - 311.

Rohrer, D., Pashler, H., & Etchegaray, J. (1998). When two memories can and cannot be retrieved concurrently. Memory & Cognition, 26, 731 – 739.

Ross, H. B., & Anderson, J. R. (1981). A test of parallel versus serial processing applied to memory retrieval. Journal of Mathematical Psychology, 24, 182 – 233.

Ruthriff, E. & Pashler, H. (submitted). Bottlenecks in Dual-Task Performance: Structural Limitation or Strategic Postponement?

Schweikert, R. (1983). Latent network theory: Scheduling of processes in sentence verification and the Stroop effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 353-383.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. Journal of Experimental Psychology: General, 117, 258-275.

Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition. Hillsdale, NJ: Erlbaum. Pp. 133-168.

Townsend, J. T., and Ashby, F. G. (1983). Stochastic modeling of elementary psychological processes. Cambridge: Cambridge U. Press.

Townsend, J. T. and Colonius, H. (1997). Parallel processing response times and the experimental determination of the stopping rule. Journal of Mathematical Psychology, 41, 392-397.

Townsend, J. T. and Nazawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories.. Journal of Mathematical Psychology, 39, 321-359.

Welford, A. T. (1952). The "psychological refractory period" and the timing of high speed performance -- A review and a theory. British Journal of Psychology, 43, 2-19.

Wenger, M. J. (1999). On the whats and hows of retrieval in the acquisition of a simple skill. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25, 1137-1160.

Footnotes

1. The finding that the parallel model prediction for RT2 is only slightly longer than RTv may be surprising. This result can be understood by noting that the difference between RT_v and RT_k is large relative to the standard deviations for those two tasks. Thus, in a race, the vocal task will usually be the second task completed. Further, when the key press task finishes second, it will usually be only

modest slower than the vocal task. The result is that the prediction for RT2 averaged over trials is only slightly slower than that for RT_v.

Table 1.
Individual subject mean RT results for Experiment 1.

Su	Task-Order	Stim-Set	RTk	RTv	RT1	IRI	RT2	Pr(voc 1st)	RT2p	RT2s
1	v,k	A	485	676	752	226	978	0.00	679	1161
2	v,k	A	696	927	1981	50	2031	.12	977	1623
3	v,k	A	849	1098	1819	45	1865	.04	1183	1947
4	v,k	A	538	805	871	91	962	.00	807	1343
5	v,k	B	633	813	1072	344	1416	1.0	844	1446
6	v,k	B	678	1142	1398	518	1917	.98	1164	1820
7	v,k	B	768	1128	869	913	1782	.00	1155	1896
8	v,k	B	871	944	2232	69	2301	.05	1131	1815
9	k,v	A	.723	883	1159	568	1724	.92	981	1606
10	k,v	A	1064	1423	2633	218	2851	.22	1580	2487
11	k,v	A	761	928	1299	570	1870	.94	1045	1689
12	k,v	A	1003	1064	1334	104	1438	.04	1180	2067
13	k,v	B	721	1255	2338	116	2454	.06	1279	1976
14	k,v	B	683	929	1271	746	2018	1.0	964	1612
15	k,v	B	623	924	1196	160	1357	.81	931	1547
16	k,v	B	545	900	896	794	1691	.05	923	1445

Notes: su = subject; stim-set = stimulus set; Pr(voc 1st) = proportion of dual task trials on which the vocal task was completed first; RT2p = prediction of parallel model for RT2; RT2s = prediction of sequential model for RT2.

Table 2.

Individual subject mean RT results for Experiment 2.

Su	Task-Order	Stim Set	RTk	RTv	RT1	IRI	RT2	Pr(voc 1st)	RT2p	RT2s
1	v,k	A	1036	1032	1351	711	2062	0.00	1382	1969
2	v,k	A	1061	1409	2534	278	2812	0.12	1649	2371
3	v,k	A	1141	1052	1888	441	2330	0.77	1430	2094
4	v,k	A	872	1124	1114	957	2072	0.04	1340	1897
5	v,k	A	685	811	1214	127	1341	0.00	979	1397
6	v,k	B	1744	2081	3190	1321	4511	0.90	2491	3726
7	v,k	B	1064	1047	1698	884	2582	0.97	1451	2011
8	v,k	B	1455	1514	2251	1118	3369	0.77	1942	2870
9	v,k	B	751	1009	1183	190	1373	0.16	1149	1661
10	v,k	B	824	1188	1598	314	1912	0.54	1339	1913
11	k,v	A	955	1115	1649	609	2259	0.66	1321	1970
12	k,v	A	1071	1145	1710	1654	3364	0.90	1478	2117
13	k,v	A	789	743	806	200	1007	0.29	993	1433
14	k,v	A	973	921	1014	663	1667	0.03	1247	1794
15	k,v	A	1222	937	1775	1089	2864	1.00	1412	2060
16	k,v	B	719	789	1036	84	1120	0.30	996	1408
17	k,v	B	757	891	1470	428	1899	0.09	1019	1548

18	k,v	B	1168	1013	1422	302	1725	0.67	1489	2082
19	k,v	B	1156	1272	2014	1283	3298	0.027	1689	2329

Notes: su = subject; stim-set = stimulus set; Pr(voc 1st) = proportion of dual task trials on which the vocal task was completed first; RT2p = prediction of parallel model for RT2; RT2s = prediction of sequential model for RT2.

Figure Captions

Figure 1. Test results of Experiment 1 averaged over all subjects.
Figure 2. Test results of Experiment 1 averaged over non-grouper subjects.
Figure 3. Test results of Experiment 2 averaged over all subjects.
Figure 4. Test results of Experiment 2 averaged over non-grouper subjects.

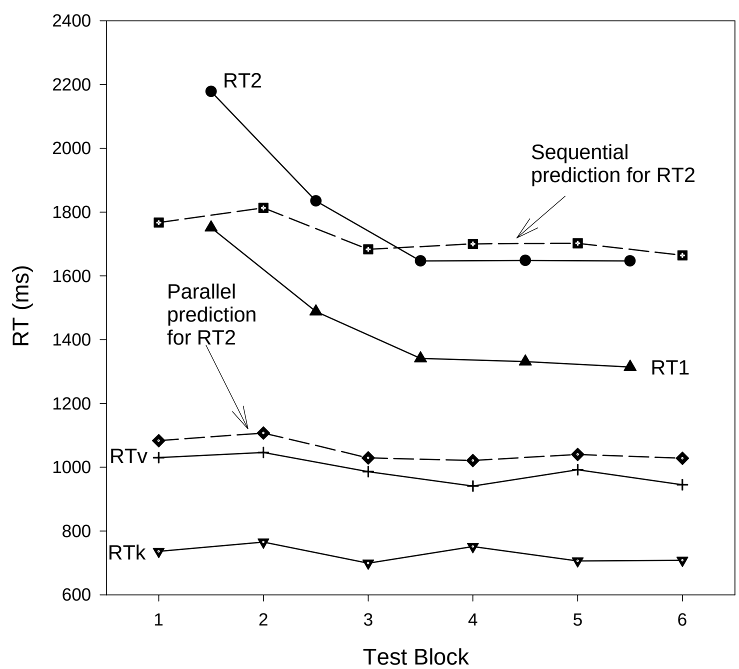
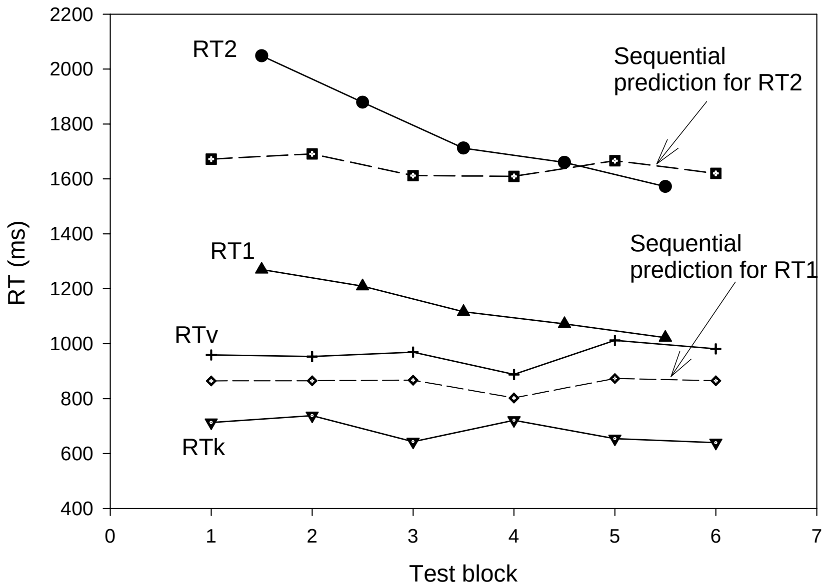
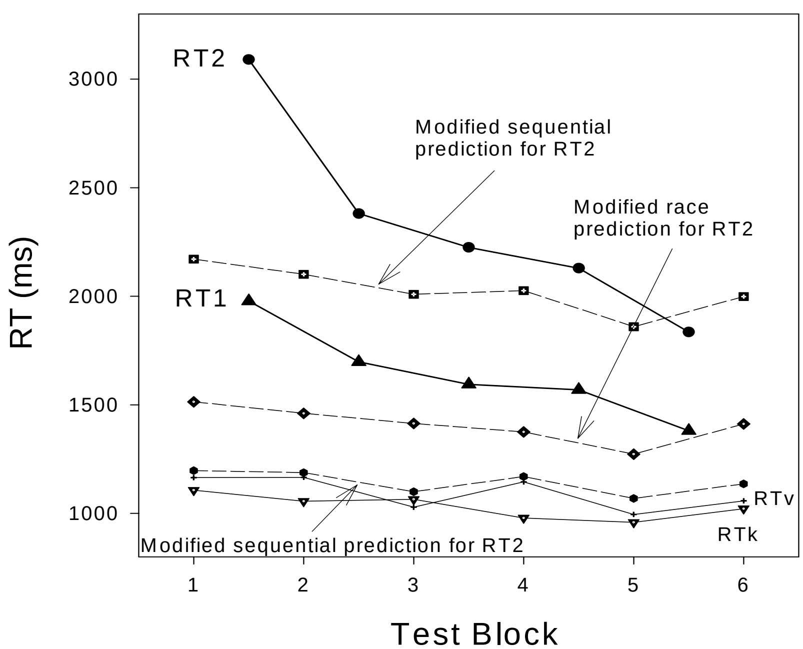


Fig. 1



RT2
Sequential prediction for RT2
RT1
Sequential prediction for RT1
RTv
RTk
RT (ms)
Test block
2200
2000
1800
1600
1400
1200
1000
800
600
400
0
1
2
3
4
5
6
7

Fig. 2



RT2
Modified sequential prediction for RT2
Modified race prediction for RT2
RT1
RTv
RTk
Modified sequential prediction for RT2
RT (ms)
Test Block
3000
2500
2000
1500
1000
1
2
3
4
5
6

Fig. 3

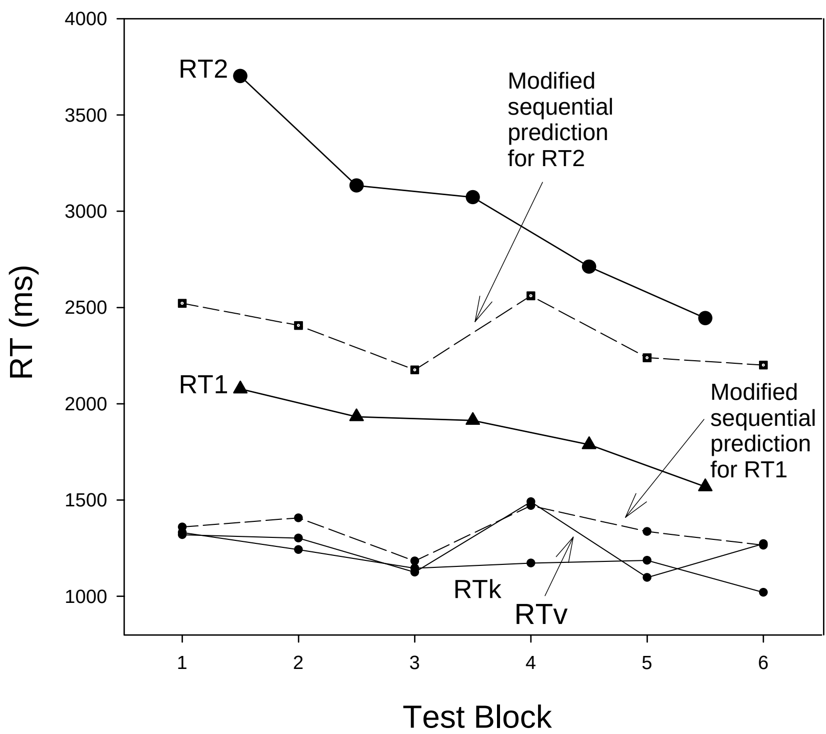

RT (ms)
4000
3500
3000
2500
2000
1500
1000
RT2
Modified sequential prediction for RT2
RT1
Modified sequential prediction for RT1
RTk
RTv
1
2
3
4
5
6
Test Block

Fig. 4