UNIVERSITY OF CALIFORNIA, SAN DIEGO

Chronometric investigations of task switching

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy in Psychology

by

Clark Fagot

Committee in charge:

Professor Harold Pashler, Chair
Professor Terry Jernigan
Professor Jeff Miller
Professor Donald Norman
Professor Benjamin Williams

1994

INTRODUCTION

In 1927 Arthur Jersild published an interesting set of experiments. Subjects in Jersild's experiments made individual responses to items in lists, sometimes performing the same task on each item of the list, and at other times alternating between performing two different tasks. For example, in his first experiment subjects were presented with lists of fifty digit-pairs. On some lists they wrote the sum of the two digits next to each pair; on other lists they wrote the product of the digits next to each pair; on yet other (alternating) lists they wrote the sum of the digits next to the first pair, the product next to the second pair, the sum next to the third pair, and so on. Jersild found that subjects were 590 ms/item slower on alternating lists than the average of the two non-alternating lists. In Jersild's first four experiments (using four different task pairs) subjects were slower on alternating lists by 590 (above), 620, 660, and 800 ms/item , respectively. Moreover, the effect remained even after moderate amounts of practice.

This alternation cost[1] indexes an aspect of performance that is rarely addressed in psychological research. Indeed, the standard approach to studying human performance and cognition is: give subjects a task, practice them on it, tell them they are about to do it, give them time to get ready to do it, present a stimulus, record a response, and finally, make inferences about what went on between the stimulus and the response. This is a perfectly sensible way to study the basic operations underlying performance on a given task. However, in the real world people do not always perform the same task over and over again. Jersild's method offers a controlled way to study the limitations of human performance when the task to be performed changes.

---

[1]The alternation cost has been called a shift cost in previous work. We opt for alternation cost because it leaves open the question of what is actually going on in the subject's mind.

There has been only a little work following up on Jersild's findings (Spector & Biederman, 1976; Allport & Styles, 1991). However, based on the research that has been done, four conclusions can be reached:

(1) When the alternation cost occurs, it is large -- a few hundred milliseconds or more per item.

(2) When the two tasks use disjoint and distinguishable stimulus sets (henceforth, <u>disjoint tasks</u>) the alternation cost vanishes or at least shrinks from hundreds down to tens of milliseconds. This was shown by Spector and Biederman (1976)[2] who had subjects subtract three from Arabic numbers and name antonyms of written words. But perhaps more convincing is one of Jersild's own experiments. In this experiment subjects cycled through <u>four</u> disjoint tasks and yet were consistently <u>faster</u> on these lists than on the pure lists.

(3) Adding to each item a cue that unambiguously informs the subject which task to perform reduces -- but does not eliminate -- the alternation cost. Spector and Biederman tacked "+3" or "-3" on to the end of each item in lists of numbers (the subject's task was to add three or subtract three), so that the stimulus plus cue informed the subject which task to do. The investigators still found a sizable alternation cost (188 ms/item). This was, however, much reduced from the 402 ms/item alternation cost found in a comparable experiment minus the appended cues.

(4) Though no alternation cost is normally found with disjoint tasks, an exception occurs when the disjoint stimulus sets for the two tasks are also the stimulus sets for another pair of tasks which the subject recently performed in alternation (Allport & Styles, 1991). Allport and Styles have called this phenomenon <u>executive</u>

---

[2]Even though some of Jersild's own data strongly suggest that disjoint tasks do not in general show an alternation cost, he believed that such a generalization was not warranted and indeed incorrect. The reasons for his belief are not worth going into here, but suffice it to say that he appears to be wrong.

proactive interference (executive PI). In Allport and Styles' (1991) Experiment 2, subjects responded to sequences of cards composed of either a single colored word (color items), or several digits (number items). Thus, subjects might have to either name the color or the word from the color items, and the digit or the number of items from the number items. In the first third of this experiment some subjects named the color of the color items and the numerosity of the number items. Meanwhile, the relevant and irrelevant attributes of the items were reversed for other subjects. As would be expected given (2), virtually no alternation cost was found in this segment of the experiment. In the next third of the experiment, subjects had to respond to the attributes that were previously irrelevant. Now significant alternation costs appeared -- averaging 250 ms on the first list after the reversal. After completing several lists with the new mappings, the alternation cost dissipated, but did not go away. A similar pattern was found in the final third of the experiment where the mappings were reversed yet again. In addition, non-alternating list performance was not affected by the reversal of the mappings.

Full Set Switch Hypothesis

Most cognitive psychologists would attribute Jersild's results to problems involving "task set" (see Spector & Biederman, 1976, for an informal survey). Task set is a well-known term in cognitive psychology, yet definitions are hard to come by. We will define a set as a state of preparation. A task set, then, is the state of preparation attained in order to perform a task. One might, then, envision a task set as the activation in memory of a set of rules that define a task. For example, the +3 task used by Spector and Biederman (1976) could be described by the following rules: if you see "1" say "4"; if you see "2" say "5"; etc. So the task set for the +3 task would correspond to having these rules activated in memory. Using this analogy we will sometimes refer to the "loading" of a task set. The alternation cost arises, according to conventional thinking, because after performing the -3 task the subject cannot perform the +3 task until the set for that task is loaded, but after this point it is no slower to add 3 to the number than it is on non-alternating lists. This hypothesis will be called the full set switch hypothesis.

It is possible, using the full set switch hypothesis, to account for the findings described in (1) through (4) above. (1) The full switch hypothesis explains the large alternation cost because subjects must complete the switch process before moving on to the next stimulus on alternating lists. (2) This alternation cost is not found when disjoint tasks are used because both sets can be held in place at once (or a combined set can be formed). That is, when each stimulus is relevant to both tasks, it is important to only have the appropriate set loaded, so task set must be changed each stimulus (hence the alternation cost), but with disjoint stimulus sets this is not a factor. (3) Adding a cue to the stimulus might reduce the alternation cost in one of two ways. First, it might make the switch occur more quickly. Second, the cue and stimulus might combine to trigger a response, thereby avoiding the need for a switch. Thus, combining the cue and stimulus allows the tasks to become disjoint. (4) To account for Executive PI, one might suppose that even with disjoint tasks, when subjects first start to alternate between two tasks they begin by alternating between the sets for the two tasks, and only after a number of stimuli do they form a combined set. In Executive PI, then, when alternation is required after the tasks change, subjects might have a tendency to switch into the "old" sets when they switch, rather than the new and appropriate ones. This is also consistent with the fact that non-alternating performance was not slowed by the task change, since in the non-alternating case subjects do not have to switch except at the start of the list.

Purpose and Basic Approach

These four findings provide clues about what is going on that allows people to control how they respond to stimuli at a given time. At least superficially, these results seem to suggest a straightforward account of task switching in terms of the need to switch task set (described in the preceding paragraph). However, at this point there is no direct support for such a view. In this paper we attempt to understand at a detailed mechanistic level why the costs associated with alternating between two tasks occur. The paper is divided into three parts. Section One is concerned with the rate at which work is done on a task on alternating lists as a function of time since the stimulus was presented. The full set switch hypothesis makes a very definite claim on this point: the slowing occurs because of the insertion of an additional process in the alternating condition, and once this has been achieved, processing occurs as fast on the alternating lists as on the pure lists. The findings of this section will test this, and will help to narrow down the set of other possible models to consider. Section Two tests the extent to which the alternation cost is a "shift cost" and the role of other various possible factors in the cost, and sketches a more detailed model. Finally, Section Three pursues questions concerning under what circumstances an alternation cost will occur with disjoint tasks.

Section One: Dynamics of processing with alternating tasks

The goal in this section is to better understand the temporal dynamics of processing when people alternate between doing two different tasks: At what rate does processing proceed during alternating list performance? Is processing on a task postponed until a switch has occurred?

Figure 1A shows the sequence of processing on non-alternating lists in the particular situation where each stimulus is presented when the response before is made. Time flows from left to right. The time-slices marked S_N and R_N represent the times when the Nth stimulus is presented and the Nth response is made, respectively. The boxes represent stages of processing. We break processing into 3 stages for reasons that will become more apparent later on. The discussion that follows will not depend on breaking processing down into precisely 3 stages or on what each of these particular stages does.

There would seem to be two extreme models of alternating list performance: On the one hand, it might be that when subjects respond to stimulus N-1, they are not yet ready to begin some stages of processing of task N. Before these stages can begin, a switch process must occur (Figure 1B). Once this has been accomplished, processing can proceed just as efficiently as on non-alternating lists. Thus, according to this account, except for the period of time in which processing on the task is being held up until the switch process is completed, processing on the alternating lists is no different from processing on the non-alternating lists, and the alternation cost is accounted for by the duration of the switch process. This model will be called the full discrete switch model[3]. On the other hand, the constant readiness model holds that while the rate of processing is slowed on alternating lists, it is not delayed by a switch

[3]Although the full discrete switch model would seem to correspond to the full set switch hypothesis, it is given a different name in this section in order to fit into this section's framework.

process (Figure 1C). Whatever the actual case, it is likely to fall somewhere on a continuum between these two models. Thus, we start by testing these models.

Notice that in Figure 1 each stimulus is presented just as the response to the previous stimulus is made. What do the models predict if stimuli are presented sometime before or after this? Consider the case in which the stimulus is presented sometime after the response to the previous stimulus. The time between the response to stimulus N-1 and the presentation of stimulus N is called the response to stimulus interval (RSI) for task N. In the case of the full discrete switch model, if the RSI is large enough that the switch process is already completed, then the alternation cost should disappear. On the other hand, the constant readiness model holds that processing will proceed at the same rate no matter when the stimulus is presented, so RSI will have no effect on RT on alternating lists, or, therefore, the alternation cost. Thus, one test of these two models is to manipulate RSI.

What if the stimulus is presented before the response to the previous stimulus (preview)? To assess this, we need to consider the effect of preview in the non-alternating condition. This topic has received a fair amount of research under the heading of serial RT (Leonard, 1953; Pashler 1994). The basic finding is that preview increases the rate of responding compared to no preview (0 ms RSI) (Cattel, 1886). Furthermore, it seems that this speedup is achieved because when preview is provided certain stages of processing can overlap (Pashler, 1994) -- namely, perception of stimulus N can overlap with response selection and execution of stimulus N-1 and response execution of stimulus N-1 can overlap with perception of stimulus N and selection of response N. When no preview is provided, however, overlap does not occur simply because stimulus N is not available for processing until response N-1 has been made.

So what do the two models of alternating list performance predict should happen when preview is provided? The full discrete switch model holds that when the stimulus for the task is presented without preview (0 ms RSI), the stages marked B in Figure 1 wait for the switch to occur rather than for completion of the stages marked A in Figure 1. Thus, although preview might allow the stages marked A to be completed earlier, the stages marked B do not begin any earlier, and thus the rate of responding is unaffected by preview. The constant readiness model, however, predicts that there will be a preview benefit in the alternating condition. This is because stages of processing for consecutive tasks should be able to overlap without slowing down just as in the non-alternating condition. The size of the preview benefit corresponds to the duration of the perceptual and response execution stages. Therefore, if we assume these stages take just as long in the alternating condition as the non-alternating condition, then it follows that the preview benefit should be the same size as in the non-alternating condition. Thus, manipulation of preview provides a further critical test of the two models.

Experiment 1

In Experiment 1 we test the constant readiness model and the full discrete switch model by manipulating RSI and preview. Subjects made button-press responses to lists of 10 colored letters presented on a computer screen from left to right. Subjects responded to either the color (red, green, or blue) or the letter (A, B, C). The experiment was divided into 18 blocks of 5 lists each. In some blocks subjects responded with the letter task on each item in a list. In other blocks they responded with the color task on each item. Finally, in still other blocks they alternated between the tasks. In addition, on some blocks stimulus N was presented immediately after the response to stimulus N-1 (0 ms RSI), while on others it was presented 1500 ms afterwards (1500 ms RSI). This latter condition should provide enough time for the switch process to occur if the full discrete switch model is accurate, since the alternation cost amounts to only several hundred milliseconds. On still other (preview) blocks two stimuli were presented at the outset, and then as each response was made an additional stimulus was displayed. This condition provided the subject with a preview of one stimulus ahead of the one to which they were currently responding. In all cases, stimuli remained on the screen even after they were responded to.

When an error was made on any of the ten stimuli in a list, the list was re-started (with new stimuli), so that if subjects made one error on each list they would never finish the experiment (they would never finish even the first list in this case). This "re-start" procedure was employed in order to keep the error rate to a minimal level. In particular, if subjects performed such that there was a 5% chance of an error on each stimulus, there would be 40% chance of re-starting the list. Thus, subjects had every incentive to keep errors to a minimum.

One additional factor was manipulated between subjects. This was whether subjects used the same set of response keys for both tasks or used a different set of keys for each. It was thought that this might influence the size of the alternation cost or even the pattern of effects.

Methods

Subjects. Thirty-six undergraduate students at the University of California, San Diego participated for partial fulfillment of a course requirement.

Apparatus and Stimuli. Stimuli were presented on IBM PC compatible microcomputers with NEC Multisync monitors, and responses were recorded on the keyboard using routines that provided millisecond accuracy. Stimuli were the capital letters "A", "B", and "C" appearing in each of the colors red, blue, and green. The letters appeared on a black background with dimension 1.2 cm in height and 0.7 cm in width, or 1.15 deg X 0.67 deg visual angle based on a typical viewing distance of 60 cm.

In the Same-Keys condition the response keys were the v, b and n keys, labeled with A-red (the letter A and a red color-patch), B-blue, and C-green, respectively. In the Different-Keys condition the color keys were as before, but the letter keys were q, a, and z for A, B, and C, respectively. So in the Different-Keys condition the keys for the two tasks were layed out perpendicular to each other, with the color-task keys oriented horizontally and the letter-task keys vertically.

Tasks. Color Task: subjects pressed the response key corresponding to the color of the stimulus. Letter Task: subjects pressed the response key corresponding to the letter of the stimulus.

Stimulus Lists. Each trial consisted of the presentation of a list of 10 stimuli. Stimuli were presented on the screen from left to right; once a stimulus was presented it remained on the screen until the end of the trial. The timing of the presentation of each the stimulus is discussed in the Procedure section. Each stimulus was offset from the preceding one by 2 cm, the first one appearing in the same location the fixation point occupied. Subjects always responded to the stimuli in left to right order. The color and letter of each stimulus was randomly determined, with the constraint that no adjacent stimuli had the same color or letter.

Design. Task Sequence and Display Condition varied between each block of 5 trials, and Key-Condition and Block Order varied between subjects. There were three possible Task Sequences: pure color, pure letter, and alternating color and letter; and three possible display conditions: preview, no preview with 0 ms RSI, and no preview with 1500 ms RSI.

There were two blocks of each of the nine possible combinations of these two factors, for a total of 18 blocks. The order of the blocks for a subset of the subjects was as follows: pure color, pure letter, and alternating color and letter all with 0 ms RSI (no preview); followed by the same sequence of the task variable with 1500 ms RSI; followed by the same sequence of the task variable with preview; finally this block ordering was repeated once more. The block ordering for other subjects was derived from the ordering above by "rotating" it 1 to 8 times, for a total of 9 possible block orderings. The Key Conditions were Same-Keys condition and Different-Keys condition. Key Condition and block order were counter-balanced across subjects (2 subjects X 2 Key-Conditions X 9 Block Orders).

Procedure. Subjects received written instructions. In addition, the experimenter re-iterated the instructions to insure that the subject clearly understood them. Subjects were told that they should make all responses as fast as possible, but that if they made an error the trial (list of 10 items) would be re-started. Subjects performed different sequences of tasks on different lists; which sequence they performed was determined by instructions displayed on the screen at the beginning of each block. The possible instructions were "Respond to colors only." (color task on each item), "Respond to letters only." (letter task on each item), and "Alternate your response between color and letter. Color first." (color task on first item, letters task on second item, color task on third item, and so on). Instructions remained on the screen for 4000 ms, and were followed by 2000 ms of blank time before the first trial of the block. There were nine practice blocks (one block per block type) with one list in each.

Each trial was initiated by the presentation of a white fixation point 6.5 cm to the left of the center of the screen for 1000 ms. Stimulus presentation began 500 ms after the offset of the fixation point. Initially, the first stimulus was presented alone, except in the preview condition in which the first two were presented. The next to-be-drawn stimulus was presented immediately after each response was made, except in the 1500 ms RSI condition, in which case it was presented 1500 ms later (i.e., after the RSI). Thus, in the preview condition subjects could always see one stimulus beyond the one they were currently responding to, but in the other conditions they could only see the stimuli up to the current one. If the subject made an incorrect response, the word "ERROR" was displayed on the screen for 1500 ms, immediately after which the trial was re-started from the beginning (with new stimuli). Trials were separated by an inter-trial interval (ITI) of 1500 ms.

Between blocks the average RT for all blocks and the number of times trials had to be re-started on each block were displayed to the subject. The subject pressed a key to continue.

Results and Discussion

Reaction Times. Mean correct reaction time (RT) is shown in Figure 2 broken down by presentation condition (0 or 1500 ms RSI or preview) and alternation condition. RT was measured from the previous response in the preview condition and from the presentation of the current stimulus in the 1500 ms RSI condition. For the 0 ms RSI/no preview condition the current stimulus was presented when the previous response was made, and this is when RT was measured from in this condition.

It turned out that when the correct response on alternating lists was the same response that would have been correct on the previous trial had the "irrelevant" stimulus dimension been the "relevant" one, subjects were on the order of 100 ms slower. Why this occurs is an interesting question, but beyond the scope of this paper. In order to facilitate the comparison between conditions we did not include these trials in the mean RT shown in the figure or used in the analyses that follow. Similarly excluded were trials in which the correct response was the same as for the previous stimulus.

Two separate analyses were carried out on the data: One for the effect of RSI (0 vs. 1500 ms RSI conditions), and another for the effect of preview (preview vs. no preview/0 ms RSI conditions).

The full discrete switch model predicts that the long RSI condition will eliminate the alternation cost, whereas the constant readiness model predicts that it will make no difference. In the 0 ms RSI condition there was a 292 ms alternation cost, raising RT from 653 ms in the non-alternating condition to 945 ms in the alternating condition. This effect was significant, $F(1,34)=280$, $p < .01$. In the long RSI condition there was a 161 ms alternation cost, raising RT from 596 ms to 758 ms. This effect was also significant, $F(1,34)=141$, $p < .01$, but it was significantly smaller than in the 0 ms RSI condition, $F(1,34) =51$, $p < .01$. Although a long RSI did reduce the alternation cost considerably, it did not eliminate it. Thus, these data are inconsistent with both the full discrete switch model and the constant readiness model.

The full discrete switch model also predicts that there will be a preview benefit in the non-alternating condition but not in the alternating condition. On the other hand, the constant readiness model predicts that there will be an equal preview benefit in both conditions. There was a 113 ms preview benefit in the non-alternating condition, and this was significant, $F(1,34)=156$, $p < .01$. In the alternating condition, there was a marginally significant benefit of preview, $F(1,34)=3.3$, $.05 < p < .10$. This is at odds with the full discrete switch model which predicts no preview benefit. However, the benefit was only 39 ms, significantly smaller than in the non-alternating condition, $F(1,34) = 11$, $p < .01$. This is inconsistent with the constant readiness model, which predicts a benefit just as large as in the non-alternating condition.

Thus, these data are inconsistent with both the full discrete switch model and the constant readiness model, but plausibly consistent with an intermediate account. One alternative to both models is that subjects are slowed on mean RT in the alternating condition merely because on some of the trials they are unsure about what task they are supposed to be doing. On this account one would expect that on most trials subjects would perform the tasks just as efficiently as on the non-alternating lists. Figure 3 shows the data for the 0 ms RSI condition broken down by alternation condition and the RT decile into which a given response fell, i.e., the data are "vincentized". This analysis reveals that the effect of alternation is found throughout the distribution, even on the very fastest trials. This refutes the hypothesis that the effect is due to only occasional slowing.

Errors. The mean number of times a list had to be restarted due to an error is shown in Table 1 as a function of display condition (preview, no preview/0 ms RSI, 1500 ms RSI) and sequence (non-alternating and alternating). The effect of display condition was not significant, $F(2,68) = 1.9$, $p > .15$, nor was the effect of sequence, $F(1,34) = 1.1$, $p > .3$. In addition, the interaction between these two factors was not significant, $F(2,68) = 2.9$, $.05 < p < .1$. Finally, key condition had no significant effects or interaction with any other variable.

Trial re-start rates allow an estimate of the per item error rates. Table 2 shows the per item error rates derived from these data[4]. These data are the estimated chance of making an error on an individual stimulus. There was no effect of display condition, $F(2,68) = 1.9$, $p > .15$ or sequence, $F(1,34) < 1$, and the interaction was also not significant, $F(2,68) = 2.6$, $.05 < p < .1$. These data show that subjects are

[4]The per item error rates are estimated with the formula:
error = 1- 10th-root of 1/(1+restarts). See Appendix for derivation.

performing at an extremely accurate level: under 2% errors, and well below what is normally seen in choice RT tasks.

Why is there a non-alternating list RSI benefit?

On the alternating lists it is natural to suppose that there is a large RSI benefit because longer RSIs allow the subject to prepare for the task they are about to do (at the expense of preparation for the other task). On non-alternating lists, however, the subject performs the same task on each item, so there is no need to prepare in this sense. Thus, it might seem odd that an RSI benefit of 57 ms was found on non-alternating lists in Experiment 1. However, an RSI benefit in serial RT (non-alternating lists) has been regularly observed in studies of serial RT (Rabbitt 1969, 1980; Wilkinson, 1990; Pashler & Baylis, 1991b). How is this benefit on non-alternating lists related to the benefit on alternating lists?

It might be that there is an interval after a response is made during which processing on a task is slowed or even postponed (a "refractory period"). Welford (1952, 1959), for example, hypothesized that the subject monitors kinesthetic feedback from the previous response, and is unable to begin selecting another response until this monitoring is complete. According to this account, the same "refractory period" should occur on alternating lists. Thus, part of the RSI benefit on alternating lists can be attributed to an effect that also occurs on non-alternating lists.

Rabbitt (1969, 1980) and Kafrey and Kahneman (1977) proposed that subjects are somehow "unprepared" for a new task early in the RSI interval, and thus cannot begin the task immediately when the RSI is short. Thus, one might suppose that early in the RSI the subjects are "unprepared" for any task, later only for the task they just did, and only still later can they become prepared for the alternate task. Notice that although the same word -- "unprepared" -- is used to account for both the alternating and non-alternating RSI benefits, this account does not give a satisfying explanation for the relationship between the two effects.

Both of the above accounts -- a "refractory period" and an "unprepared" period -- are equivalent at a more abstract level. They both suppose that there is a period after a response has been made during which, on both alternating and non-alternating lists, processing does not occur (or more generally is slowed). After that, processing on alternating lists is further delayed or slowed.

This type of an account, however, is not necessarily correct. The faster responding at long RSI in serial RT is usually accompanied by a higher error rate (Krueger and Shapiro, 1981; Wilkinson, 1990; Pashler and Baylis, 1991B; but see Rabbitt, 1980, Experiment 2). Thus, it is possible that the entire speed-up is explained by a speed-accuracy tradeoff. On the non-alternating lists in Experiment 1 there was only a small trend toward more errors at short RSI (0.1%, $F < 1$). However, at the very high levels of accuracy found in this experiment (over 98%) a small effect on accuracy could correspond to a large effect on RT.

Assuming, for the moment, that a speed-accuracy tradeoff as found in serial RT is also present on our non-alternating lists (using the trial re-start procedure), the question then becomes whether the same speed-accuracy tradeoff occurs on alternating lists. If it does, then some of the RSI benefit on alternating lists might reflect a related problem on non-alternating lists. Furthermore, it might even be the case that the entire RSI benefit on alternating lists is caused by such a speed-accuracy tradeoff. However, on alternating lists in Experiment 1 there were fewer errors at the long RSI (1.6% compared to 2.0%, $F(1,34) = 3.0$, $.05 < p < .1$). This suggests that the speed-accuracy tradeoff that occurs in serial RT, and possibly on the non-alternating lists in Experiment 1, does not occur on the alternating lists. Thus the RSI benefit on alternating lists might have a completely different cause.

Wilkinson (1990) has suggested a particular hypothesis of how the speed-accuracy tradeoff over RSI occurs in serial RT. His suggestion is that there is a "partial refractory period" that is "biased" toward response initiation. That is, there is a tendency to avoid making responses close together in time, and this is independent of the amount of information that has accrued for the response. However, what he called the sensory/perceptual/decision-making stages (SPD stages) are less affected by RSI. Thus, at short RSI the SPD stages proceed just as at long RSI, but since response initiation tends to occur later, they operate longer and arrive at the correct response more often. Consistent with this hypothesis, the RSI benefit is reduced when the number of response alternatives is increased (Wilkinson, 1990).

This account explains both why there would be a speed-accuracy tradeoff on non-alternating lists and why there would not be one on alternating lists. That is, on alternating lists a response is never ready to be initiated early enough for it to be affected by the "partial refractory period". However, the data do not speak to whether a speed-accuracy tradeoff accounts for the entire RSI benefit in serial RT. Thus, there still might be a common cause behind the RSI benefits on alternating and non-alternating lists, such as the ones offered above.

In summary, one might have thought at the outset that whatever causes the modest benefit of RSI on non-alternating lists would also be having an effect on alternating lists, and possibly even that the RSI benefit on alternating lists was an enlarged version of the same effect. However, the apparent absence of a speed-accuracy tradeoff in the alternating condition suggests that the RSI benefit that occurs on non-alternating lists does not contribute to the benefit on alternating lists. It is unclear whether there is also a common contributing factor such as a "refractory" or "unprepared" period.

A more general framework: The readiness function

The results of Experiment 1 show that neither of the two basic models considered are consistent with the data. Thus, we need to consider a larger class of models of the dynamics of processing on alternating lists. For this purpose, we introduce the idea of a readiness function. Define readiness for a task as the rate at which work can be done on the task at that time. It is assumed that each task requires a fixed amount of work to be completed before a response can be made. The readiness function is a non-negative real-valued function of time t, where t=0 is assumed to be the time at which the previous response was made and the height of the function represents readiness for the upcoming task at time t (larger numbers correspond to more readiness). Thus, the function shown in Fig4A represents the full discrete switch model and the function in Figure 4B the constant readiness model (in both cases the dotted line represents the fixed level of readiness in the non-alternating condition)[5].

Intermediate models can also be represented in this way. For example, Figure 4C shows the partial discrete switch model. The only difference between this and the full discrete switch model is that readiness does not rise all the way to non-alternating levels in this case, so even if the subject is allowed a long time to prepare for the next task, responses will not be as quick as on non-alternating lists. Finally, Figure 4D shows the gradual switch model, in which the subject gradually becomes more and more ready for the upcoming task as time progresses, readiness maybe or maybe not eventually rising to non-alternating list levels (the later being the case in the figure).

---

[5]If there was a "refractory" or "unprepared" period on non-alternating lists, as suggested might be the case in the paragraph above, then readiness on non-alternating lists would not be constant as is suggested by the dotted line. However, this will not affect our arguments, and so we will not pursue it further.

This hypothetical readiness function can be partially recovered from the function relating RT to RSI. Since we assume that the amount of work that must be done before a response is made is constant, say k, we can write:

$$\int_{RSI}^{RT+RSI} r(t)dt = k,$$

so,

$$\frac{d}{dRSI}\int_{RSI}^{RT+RSI} r(t)dt = 0,$$

and by the fundamental theorem of calculus,

$$(\frac{dRT}{dRSI}+1)\cdot r(RT+RSI) - r(RSI) = 0,$$

and it follows that,

$$\frac{dRT}{dRSI} = \frac{r(RSI)}{r(RT+RSI)} - 1.$$

Therefore, the slope of the function relating RT to RSI reflects the rate at which work can be achieved at the end of the RSI compared to the rate at which work can be achieved at the time at which the response is made (i.e., at RT+RSI). For convenience we will write the slope of RT as a function of RSI as RT'.

One difficulty is that one cannot assume that the readiness function is fixed from trial to trial. For example, in the discrete switch models, the durations of the switch process might vary from trial to trial. Thus, we allow that the readiness function might change from trial to trial, and represent the readiness function on trial i as $r_i(t)$. Nevertheless, even if the readiness function arbitrarily changes from trial to trial, the following assertions follow (here we assume that $r_i(t)$ is non-decreasing):

(a) $r_i(t)=c$ for all t>A and all trials i if and only if, RT' = 0, for RSI > A.

So if the readiness function is flat for $t > A$ on every trial, then mean RT as a function of RSI is flat for RSI > A, and visa versa.

(b) $r_i(t) = 0$ for all $t < A$ and all trials i (i.e., task processing always waits until after time A) if and only if, RT' = -1 for all RSI < A

So if on no trial does processing begin before time A, then the slope of RT as a function of RSI is -1 for RSI < A. This makes sense: presenting the stimulus a moment earlier will only add an equal amount of time to RT since no work will be done during that moment.

(c) RT' is bounded by -1 and 0, and larger (less negative) values of RT' correspond to larger values of $Er_i(t)$, where the expected value is over trials.

So RT' is monotonically related to readiness. However, the exact relationship is not determined.

One important aspect to the readiness function not addressed above is the eventual state of readiness for the task after a long RSI (after RT' goes to zero). This can be assessed by whether RT approaches non-alternating list levels with long RSI's. If even after a long RSI the RT's are slower than on non-alternating lists, it follows that readiness does not reach the non-alternating list level.

Let us consider the discrete switch models (both full and partial) in more detail. A defining attribute of these models is that the readiness function is at zero for an extended duration early on. Due to b), this corresponds to a -1 slope of RT as a function of RSI. Since $r_i(t)$ might vary from trial to trial, the switch (the interval during which $r_i(t)=0$) might sometimes be very short. In this case, we would not expect a -1 slope even though a discrete switch model might be the case. One way to overcome this problem is to use small values of RSI. An additional problem is that the switch might actually begin before the response to the previous task is made. Thus if the switch on some trials only takes 100 ms and if it begins 100 ms before the response is made, a -1 slope would not be found at all with positive RSI's. This is where preview is important. For our purposes, preview is a negative RSI. If no work on task N can be achieved around the time of response N-1, then no preview benefit will occur.

To summarize, then, defining the readiness function allows us to consider a larger class of models. The function relating RT to RSI gives us information about the underlying readiness function. The RT function will have a -1 slope when readiness is at 0, or, in other words, task processing is postponed. It will have a slope of 0 when readiness is no longer changing with time. The level of readiness near the beginning of the RSI can be assessed by manipulating preview. In particular, if no work is completed on the next task until somewhat after the response to the previous task is made then there should be no preview benefit. On the other hand, if some work is done before the response a preview benefit is expected. The eventual level of readiness achieved after a long RSI can be assessed by comparing RT on alternating lists with long RSI's to RT on non-alternating lists.

Experiment 1 shows three things about the readiness function. First, since a long RSI did not eliminate the alternation cost, the eventual level of readiness would seem to be lower than on the non-alternating lists. Second, since the preview benefit was drastically reduced over non-alternating condition, work on task N must be done less efficiently at the start of the RSI interval than just prior to response N. Third, if the marginally significant effect of preview is real, either the discrete switch models are wrong, or the switch to task N is sometimes through shortly after response N-1 is made.

The following experiment samples more RSI's to plot the readiness function in more detail. This will help us see how quickly readiness reaches its eventual level on alternating lists.

Experiment 2

In Experiment 2 RSI's of 0, 200, 400, 800, and 1500 ms were used. The experiment was broken down into blocks of 10 lists. On eight of these lists per block the RSI was randomly determined separately for each stimulus from the set of five listed above. For the other two lists in each block RSI was fixed at 0 ms for one and 1500 ms for the other, as in Experiment 1.

Method

The methods used in this experiment were the same as in Experiment 1, except as noted.

Subjects. Sixteen subjects participated for partial fulfillment of a course requirement.

Apparatus and Stimuli. The stimuli in this experiment were colored shapes. The colors were red, green, blue, and yellow; the shapes were a circle, a pie shape, a square, and an X. The response keys were v, b, n, and m in the Same-Keys condition, and these plus 1, q, a, and z in the Different-Keys condition, and were labeled with the colors red, green, blue, and yellow, respectively, and the shapes circle, pie, square, and X, respectively. The shapes were all 1 cm by 1 cm and the "X" was 1.2 cm high and

0.7 cm wide. In the Different-Keys condition, the color task keys were oriented horizontally and the shape task keys vertically.

Tasks. Color Task: same as Experiment 1, extended to include the additional color. Shape Task: subjects pressed the response key corresponding to the shape of the stimulus.

Stimulus Lists. The colors and shapes were randomly determined in this experiment, with the following constraints. If a stimulus was red or a circle, the next stimulus could be neither red nor a circle. Similar constraints held for green and pie shapes, blue and squares, and yellow and X's. These constraints insured that in the Same-Keys condition, the response key associated with the "irrelevant" attribute would not be associated with either attribute of the succeeding stimulus (see Results Section of Experiment 1).

Design. Experiment 2 was broken down into twelve blocks of ten trials each. Task Sequence and RSI were manipulated. There were four possible Task Sequences: pure color, pure shape, color-shape (alternating, color first), and shape-color (alternating, shape first). The RSI's used were 0, 200, 800, and 1500 ms. Task sequence varied across blocks, and the ordering was determined analogous to Experiment 1 (for a total of 4 possible block orderings). RSI varied within each list of stimuli, so that the RSI between each pair of stimuli was randomly drawn from the four possible values. The exception to this was that on two lists in each block the RSI was fixed throughout the list: at 0 ms for one of them and 1500 ms for the other. These will be called No Vary lists. The first No Vary list in each block was randomly chosen from one of list 1 through 5, and the second No Vary list was 5 lists later. Whether the first No Vary list was 0 or 1500 ms RSI was randomly determined for the first block. The RSI of the first No Vary list in succeeding blocks alternated between 0 and 1500 ms. Finally, Key Condition was manipulated between subjects as in Experiment 1.

Procedure. As in Experiment 1, instructions were displayed to subjects at the beginning of each block telling them what task sequence to perform. In this experiment, the format of the instructions was simplified slightly. The instructions were "Color Task", "Shape Task", "Color Task -- Shape Task", or "Shape Task -- Color Task". They were displayed for 2000 ms, rather than 4000 ms as in Experiment 1. In addition, the same instructions were displayed just under the fixation point at the start of each trial (in Experiment 1 subjects saw the instructions only at the beginning of each block).

Results and Discussion

Reaction Times. Mean correct RT is shown in Figure 5 broken down by RSI, alternation, and whether RSI varied within a list or was fixed. Whether subjects did the task on the same or different sets of keys had no significant effects or interactions with any other variable, so the data are averaged across key condition in the figure and the following analyses.

The data are similar to the previous experiment for RSIs of 0 and 1500 ms. The alternation cost at RSI=0 ms was 314 ms, $F(1,14)=223$, $p < .01$, whereas the alternation cost at RSI=1500 ms was only 201 ms, $F(1,14)=41$, $p< .01$, and this difference was significant, $F(1,14)=39$, $p < .01$. When RSI was fixed within a list subjects were on average 14 ms faster than when RSI varied, and this effect was marginally significant, $F(1,14) = 3.1$, $.05 < p < .1$. This slight speedup was almost identical in both the alternating (13 ms) and non-alternating conditions (15 ms), $F <1$.

One might argue that the reason long RSI did not eliminate the alternation cost in Experiment 1 was that subjects do not need much time to perform the switch and (however paradoxically) put off beginning the switch until near the end of the RSI, often putting it off too long and incurring an unnecessary alternation cost as a result. In Experiment 2 RSI varied within lists and took on intermediate values between 0 and 1500 ms. Therefore, subjects did not know how long they had before the stimulus would appear. Thus, subjects should have had more incentive to perform the switch right away. Nonetheless, a large alternation cost was still found at an RSI of 1500 ms. In addition, two of ten lists in each block had fixed RSI's of 0 or 1500 ms for each stimulus. Thus, if there was any tendency among the subjects to put off performing the switch right away, this tendency should have been larger on the fixed RSI lists, and the alternation cost should have been greater at 1500 ms fixed RSI than 1500 ms varied RSI. This did not occur. Thus, it seems unlikely to us that any of the residual alternation cost at 1500 ms is a result of any "laziness" on the part of the subjects.

For both alternating and non-alternating lists, the most dramatic effects of RSI occurred within a 200 ms RSI. In the non-alternating case, in fact, there was a 34 ms benefit of 200 ms RSI over 0 ms RSI, but no additional benefit after 200 ms. In the alternating case, there was a 71 ms benefit for 200 ms RSI over 0 ms RSI, but only an additional 40 ms and 38 ms benefit for RSIs of 800 and 1500 ms, respectively. These numbers can be converted into slopes of RT as a function of RSI. In the alternating condition, the slope between 0 and 200 ms RSI was -.36, between 200 and 800 ms was -.07, and between 800 and 1500 ms RSI was -.05. In terms of the readiness function, this suggests that readiness is nearly at its asymptotic level after an RSI of only 200 ms.

Recall that the slope of RT as a function of RSI reflects the ratio of readiness when the stimulus is presented to readiness when the response is made. Since readiness appears to reach an asymptote very quickly, by the time a response is made readiness is almost always near asymptotic levels. It follows that RT'+1 is a rough approximation to the ratio of the average readiness at a given RSI to asymptotic readiness. This is plotted in Figure 6.

Errors. The by-item error rates for Experiment 2 were 1.6% and 1.9% for non-alternating and alternating lists, respectively. These numbers correspond to an average of 0.273 and 0.332 restarts per list, respectively. As in Experiment 1, the by-item error rate is very low, below 2% here. The effect of alternation on the by-item error rate was marginally significant, $F(1,14) = 4.5$, $.05 < p < .1$. There was no significant effect of key condition $F<1$, or interaction of key condition and alternation, $F(1,14) = 2.1$, $p > .15$.

Two components of the alternation cost

These results allow us to distinguish two components to the alternation cost: one that is overcome during the RSI (the RSI component) and one that is only slowly or not at all overcome during the RSI (the baseline component). The alternative, that there is a single component which can be overcome with a large enough RSI is not strictly ruled out, since RT might not have reached an asymptote yet in our data. One way to test between these two alternatives would be to find the asymptote. However, this does not seem practical to us since one consequence of using longer RSI's is that fewer trials can be collected in the same time period. Moreover, it is probably the case that motivational and/or alertness factors would contribute to longer RTs as RSI increased much beyond a second on both alternating and non-alternating lists. In any case, even if very long RSI's eliminate the alternation cost, there is still clearly a period in which readiness increases rapidly (before 200 ms) and a period in which it increases only slowly (after 200 ms). It is possible that these periods are affected by two different sets of factors.

Assuming this two-component hypothesis is correct, it is worth considering in more detail the nature of the RSI component. In particular, what form does the readiness function take on individual trials? It might, for example, resemble the average readiness function shown in Figure 6. On the other hand, the RSI component may reflect a discrete stage that must be completed before processing on the next task can begin. I.e., the RSI component might be a switch process (Figure 1C).

If the RSI component reflects a discrete switch process then the function relating RT to RSI informs us of its distribution. It is possible to derive exactly the distribution of a discrete switch process from this information, but it requires more reliable data on the single subject level than we have. We can, however, get a rough picture of the variability of this hypothetical process from the data we do have. Earlier it was argued that when a switch process always takes longer than a particular RSI, RT as a function of RSI will have a -1 slope at that RSI. In Experiment 2 the smallest pair of RSI's were 0 and 200 ms, and the slope between these values, was -.36. Thus, if the RSI component reflects a discrete switch it would on some trials take less than 200 ms. The same logic also asserts that there should be a preview benefit on alternating lists only if the switch is sometimes completed before the response to the previous task (which is possible only if the switch begins before this also). Thus, if the marginally significant preview benefit found in Experiment 1 is real, then the switch would have to sometimes be very quick indeed. On the other hand, since RSI continues to help even up to RSI's of 1500 ms, the switch process would sometimes take very long. Thus, if the RSI component reflects a discrete switch process it is highly variable in duration. The average speedup that is achieved by providing a large RSI reflects the average amount of time that the alternation cost adds to overall RT. The average duration of the switch process is then this speedup plus however long before response N-1 that the switch to task N begins. Thus, the RSI benefit of 149 ms at 1500 ms RSI would indicate that the switch has an average duration of at least 149 ms, possibly around 250 ms.

Summary

We asked the question of how readiness changes as one prepares for the upcoming task in the alternating condition of Jersild's paradigm. The results show: 1) readiness increases rapidly up to 200 ms RSI and more slowly thereafter, 2) readiness never reaches levels attained on non-alternating lists (at least without RSI's much longer than 1.5 seconds), and 3) there may be two components to the alternation cost, one overcome during the RSI (RSI component) and one not (baseline component). 4) If the RSI component reflects a discrete switch process, then the duration of the switch is at least 150 ms on average, sometimes completed before 200 ms RSI (and maybe before 0 ms RSI) and sometimes only after 800 ms RSI.

Section Two: A functional analysis of alternating tasks

The observations of Section One set the stage for an attempt to understand the functional workings of task switching. These results suggest that the alternation cost can be broken down into two components: one that is overcome during the RSI -- the RSI component -- and one that is not -- the baseline component. We will start by exploring some of the reasons why, in principle, the alternation cost might arise. This will allow us to formulate a taxonomy of the various possible reasons why the cost arises. We will then consider the relationship between this taxonomy and the two components suggested by the results of Section One. These logical considerations are then applied in Experiments 3 and 4. The results of these experiments have important implications and lead us to explore a number of more specific models in the remainder of Section Two.

Why does the alternation cost occur?

We begin by considering five possible reasons why the alternation cost occurs. Some of these reasons can only explain the baseline component whereas some can only explain the RSI component. Some could explain either or both. Below, we will consider the relationship of each of the following accounts to the effects of RSI. Throughout the next few pages several new terms and concepts will be introduced. To aid the reader in following these new terms and concepts we have provided Table 3.

The first three of the accounts of the alternation cost (part or all) assert that when the stimulus is presented on alternating lists subjects are prepared for the wrong task. In the Introduction we described the full set switch hypothesis which accounts for the alternation cost by supposing that a set switch process must act before each response can be selected on alternating lists, and this process puts subjects in a state of preparation equivalent to the state they are in on non-alternating lists. This model is ruled out by the existence of the baseline component. This does not, however, rule out the idea that being prepared for the wrong task when each stimulus is presented accounts for some of the alternation cost.

1. Set switch. One might suppose that there is a partial set switch that occurs during the RSI that puts the subject in a state of preparation for one task over the other but not to the extent found on non-alternating lists. This set switch could be a discrete switch (Figure 4C), so that it must be completed before selection of the response can begin. Or, it could be a gradual switch (Figure 4D), such that selection of the response begins once the stimulus is identified, but proceeds more slowly because the subject at this time is still inadequately prepared for the task.

2. "Tuning" effect. Some aspects of preparing for a task, on the other hand, might not be accomplished with a set switch that can occur during "free time", i.e. during the RSI. Instead, one might need to do the task once or twice in a row before being fully prepared for it. This type of an effect would seem to naturally follow from the notion that the machinery that selects responses becomes better "tuned" for one task or the other when doing the task (hence the term "tuning" effect).

3. Set decision. Another reason subjects might be slowed on alternating lists is that they must decide which of the two tasks they are supposed to do on each stimulus, and this decision might take time. On non-alternating lists, on the other hand, the same task is done on each stimulus so it is possible that no such decision must be made.

The potential causes of alternating list slowing cited above explain the alternation cost in terms of the need to prepare for the upcoming task. It might be, however, that the alternation cost is caused by activation of the other task set. On non-alternating lists only one of the two tasks is performed on that list, so the other task set can be suppressed as much as possible. On alternating lists, on the other hand, both tasks are performed on each list, so it is reasonable to assume that neither task can be fully suppressed at any point on the list. If this view is correct, then one might suppose that part of the alternation cost arises because there are two activated mappings relevant to each stimulus on alternating lists but only one such mapping on non-alternating lists. Indeed, as pointed out in the Introduction, there is no or little alternation cost if disjoint tasks are used (e.g., colored disks and white letters when the tasks are a color task and a letter task). In this case each stimulus is relevant to only one activated mapping -- each stimulus, in fact, is relevant to only one mapping from the whole experiment.

4. Task competition. It might be, then, that part or all of the alternation cost arises due to the rules for both tasks being activated, and not due to a need to switch sets. We will call any cost that depends on the current stimulus being relevant to two tasks a task competition cost. This distinguishes task competition effects from set switching effects since the later would occur even if the stimulus was only relevant to one of the tasks. A particular form of task competition that has been observed in many different situations is response competition. According to a response competition model the correct response for each activated task set is itself activated, and when two different responses are activated it takes longer to select the correct one. A classic example of response competition is the Stroop task (Stroop, 1935). In the Stroop task subjects are approximately 100 ms slower to name the color of a color-word when the color-name does not match the color of the word (incompatible condition) than when it does (compatible condition).

5. Criterion effects. Finally, it might be that subjects are slower on alternating lists partly because of a criterion effect. That is, subjects might adopt a more

conservative criterion for selecting responses on alternating lists than on non-alternating lists. The trial restart procedure did an excellent job of keeping errors to a minimum in all conditions of Experiment 1 and 2. However, subjects might still be more susceptible to errors on alternating lists. Thus, although in both Experiment 1 and 2 the by-item error rate was higher on alternating lists subjects still may have been selecting responses more conservatively -- and hence more slowly -- in the alternating condition. Note that since alternating lists tend to have higher error rates than non-alternating lists, we can eliminate a "pure" speed-accuracy tradeoff, where processing is just as efficient on both lists but subjects just choose a stricter criterion on alternating lists.

Each of the potential causes of slowing discussed above could potentially contribute to the alternation cost. It is not necessarily the case, however, that these causes are independent sources of slowing. For example, although a criterion effect may, on the surface, be the reason that some of the slowing occurs, the reason that subjects are more susceptible to errors on alternating lists must also be considered. For example, this reason may be task competition. Moreover, task competition might (but not necessarily) be modulated by the extent to which subjects are prepared for the upcoming task when they first get the stimulus. Thus, these different potential causes of slowing might be merely different manifestations of the same underlying problem. However, it is still worth considering them separately since they are potentially independent causes of slowing. Moreover, even if they are just different manifestations of the same problem we will soon see that we can separately remove some of them.

Mixed list cost vs. shifting cost

As noted above, one difficulty subjects face on alternating lists that they do not face on non-alternating lists is that they must "keep in mind" two tasks, since they will soon have to do the other one. Another difficulty is that the task they perform on each stimulus of the list is different from the task they performed just before. It would be useful when trying to distinguishing the potential accounts described above to have a way to assess the cost associated with each of these two difficulties separately.

Consider the following task sequence: A-A-B-B-A-A-B-B ... We will call this sequence the AABB sequence. Analogously, we will sometimes refer to a non-alternating sequence as an AAAA sequence and an alternating sequence as an ABAB sequence. When task A is performed after task B in this sequence (the response to an AABB-different stimulus) an alternation cost just like the one found on alternating lists would be expected, because as on alternating lists subjects performed a different task from just before. However, performing task A after task A (the response to an AABB-same stimuli) is a different matter, since there is no change in the task being performed. Any cost in this case compared to non-alternating lists is attributable to the fact that there is more than one task on the list. AABB lists, like alternating lists, are mixed lists because they have more than one task on them. This is in contrast to a pure list, which consists of only one task performed over and over (i.e., a non-alternating list). We will call the difference between RT to AABB-same stimuli and RT to stimuli on AAAA lists a mixed list cost. AABB lists would seem to partition the alternation cost into two parts: a mixed list cost (the difference in RT between AABB-same and stimuli from pure lists) and a shifting cost (the difference in RT between AABB-different stimuli and AABB-same stimuli). The former reflects the difficulties associated with responding to stimuli when more than one task relevant to the stimulus is activated; the latter reflects the difficulties specific to having done a different task on the preceding stimulus.

Putting it all together

So far we have introduced two ways of breaking the alternation cost into parts -- the RSI and baseline components and the mixed list and shifting costs -- as well as five potential causes for the slowing on alternating lists. How is all of this related?

First of all, the break-down of the alternating cost into the RSI and baseline components is orthogonal to the break-down into the mixed list cost and the shifting cost: It is possible that either the mixed list cost or the shifting cost corresponds to the RSI component and the other to the baseline component. It is also possible that both of these costs have baseline and RSI components, or that one of them accounts for both the baseline and RSI components and the other is not actually present (or has zero cost).

Table 4 shows the relationship of the various potential causes of slowing discussed above onto the 2 X 2 break-down of the alternation cost. Some costs can only be found in particular components. Set switching and "tuning" costs are confined to the RSI and baseline components of the shifting cost, respectively. The set decision cost, on the other hand, would be part of the RSI component since a decision of what task to perform can be made without the stimulus. Whether it is part of the mixed list cost or the shifting cost, however, is uncertain. One might imagine that on mixed lists the subject must make a decision about what task to do for each stimulus in the list. In this case the set decision cost would be in the mixed list cost. It might be, however, that on AABB lists the subject can one task twice in a row without making a separate decision to do the second task. If this were true, then the set decision cost would be a shifting cost, since the decision would be made on AABB-different but not AABB-same stimuli. Even if this were true, if the task sequence were more complicated, say ABABBABAAB, then it might turn out that the subject would need to make the decision for each stimulus. Thus, the set decision cost maps onto Table 4 in a very complicated manner, and might even depend on the particular task sequence used to assess the mixed list versus shifting cost.

We have no basis to limit the effects of task and response competition. The size of the effects might depend on the RSI, but they might not; they might depend on the previous task, but they might not. Thus, task and response competition could affect any combination of the four parts of the alternation cost.

Since task and response competition may affect how susceptible to errors the subject is and this, in turn, is the controlling factor behind criterion effects, a criterion effect can affect any component that task and response competition can affect. However, consider what criterion effects on the shifting cost are. Criterion effects on the shifting cost would mean that subjects are more susceptible to errors after doing the other task on the stimulus before, and thus they adjust their criterion accordingly. If during the RSI subjects becomes less susceptible to errors, hence re-adjusting their criterion and selecting responses more quickly, then the effect will be on the RSI component of the shifting cost. But notice, then, that this is just another form of a switch process in which what is accomplished is that the susceptibility to errors is reduced and the criterion is adjusted accordingly. If the subject does not become less susceptible to errors during the RSI, then the criterion effect is on the baseline component of the shifting cost. But then the criterion effect can be equated with a "tuning" effect, i.e., doing a task makes one less susceptible to errors when doing it again, and responses can therefore be selected less conservatively, hence more quickly. Thus, although criterion effects can occur on all four components, we will only consider them separately on the baseline and RSI components of the mixed list cost.

There are a large number of ways that the possibilities we have presented can combine to account for the cost people incur when alternating between two tasks. In

Experiment 3 we estimate the size of each of the 4 components shown in Table 4. We will then consider response competition models. The data from Experiments 1 and 3 will allow us to determine the role of response competition in the alternation cost. In Experiment 4 we test the more general class of task competition models. These experiments will lead to a much simpler view of the alternation cost. But first, we will consider a point of methodology.

The locus of factor effects

In the following experiments we will want to know which of the four parts of the alternation cost are affected by particular manipulations. Before proceeding, it will prove useful to consider how to do this in general. Suppose a factor (for example, response competition) affects the size of the alternation cost. We saw above how to divide the alternation cost into the mixed list cost and the shifting cost. This allows us a way to assess which of these costs a factor effects (if not both). But how do we assess whether a factor affects the baseline or RSI component?

Which of the baseline and RSI components a given factor affects can be determined by manipulating RSI. The critical observation is that at long RSI the RSI component does not contribute to RT. Thus, if a factor has its effect on the RSI component, then at long RSI the effect of the factor should vanish. On the other hand, if a factor has its effect on the baseline component then the effect of this factor should not be changed by RSI. It might also be that a factor affects both components of the alternation cost. In this case the factor should interact with RSI because, once again, the RSI component contributes to RT at short RSI but not long RSI. However, the effect of the factor should still be present at the long RSI since the effect of the factor on the baseline component is still relevant here.

Furthermore, this same logic can be used to separate the mixed list and shifting costs into their RSI and baseline components: at long RSI only the baseline component of the mixed list cost (shifting cost) should be present. If the mixed list cost (shifting cost) is smaller at longer RSI than short RSI, then it has an RSI component.

Experiment 3

In Experiment 3 we divide the alternation cost into the four components discussed above. We used an AABB sequence in addition to the AAAA and ABAB sequences used before. In addition, RSI's of 0 and 400 ms were employed in order to determine whether the shifting cost is part of the RSI component or part of the baseline component, or both, and similarly for the mixed list cost.

Method

The methods used in this experiment were the same as in Experiment 2, except as noted.

Subjects. Twenty students participated for partial fulfillment of a course requirement.

Apparatus and Stimuli. The stimuli were colored letters. The colors were the same as those used in Experiment 2; the letters were Q, V, L, and T. The response keys were labeled as in Experiment 2 except that Q, V, L, and T took the place of circle, pie, square, and X, respectively.

Tasks. Color Task: Subjects pressed the response key corresponding to the stimulus color. Letter Task: Subjects pressed the response key corresponding to the stimulus letter.

Stimulus Lists. The colors and letters that occurred on each list were generated analogously to Experiment 2, including the constraints placed in order to insure that the response to the "irrelevant" attribute of one stimulus could not be the correct response to the next stimulus.

Design. There were 18 blocks of 5 lists each, with each list containing 10 items. Task sequence was manipulated between blocks. There were six possible task sequences: pure color, pure letter, color-letter, letter-color, color-color-letter-letter (two colors in a row, followed by two letters, followed by two colors, etc.), and letter-letter-color-color. The order of the blocks was determined as follows: the first three blocks were some random permutation of pure color, color-letter, color-color-letter-letter or some permutation of the other three block types. The next three blocks (blocks 4-6) were a permutation of the three blocks types not used in the first three blocks. Finally, this order of the first six blocks was repeated two more times for a total of 18 blocks.

RSI and key condition were also varied. RSI was either 0 or 400 ms and was randomly determined for each stimulus. Key Condition varied between subjects as before.

Results and Discussion

Reaction Times. Figure 7 shows correct RT to stimuli from AAAA lists, stimuli from ABAB lists, AABB-same stimuli, and AABB-different stimuli as a function of RSI. On ABAB lists subjects might receive a red stimulus for the color task and then two stimuli later receive a red stimulus for the color task once again. It turns out that subjects are much faster on these fortuitous trials. On AABB-different stimuli, however, the task and stimulus can be repeated only in three stimuli, and on AABB-same stimuli only in four. Thus, in order to facilitate comparison between conditions, trials in which the task and relevant dimension was repeated on one of the three previous stimuli were not included in the analysis (or in the figure).

The results from the AAAA and ABAB lists resemble previous findings of the effect of RSI on the alternation cost. There was a 441 ms alternation cost, and this was significant, $F(1,18) = 220$, $p < .01$. RSI had an overall significant benefit of 86 ms, $F(1,18) = 123$, $p < .01$; a 70 ms/item benefit on AAAA lists and a 122 ms/item benefit on ABAB lists, and this difference was significant, $F(1,18) = 6.8$, $p < .05$.

There was one respect in which the data from AAAA and ABAB lists differed from previous experiments: Key condition (same keys vs. different keys) significantly interacted with alternation, $F(1,18) = 6.6$, $p < .05$. The alternation costs were 523 ms and 364 ms for same and different key conditions, respectively. Although there was a small trend in the same direction in previous experiments in no other experiment was the difference nearly so large (55 and 18 ms differences in Experiments 1 and 2, respectively, compared to 159 ms here). It is unclear whether this difference arose due to the presence of AABB lists or just by chance[6]. In a pilot experiment with only AABB lists and AAAA lists there was an 84 ms difference between the alternation costs (calculated by mean RT on AABB-different stimuli minus mean RT on AAAA lists) in the same keys and the different keys conditions, which is smaller than found here but larger than in other experiments.

There was both a shifting cost and a mixed list cost in this experiment. The shifting cost (AABB-different minus AABB-same) was 197 ms, $F(1,18) = 90$, $p < .01$, and the mixed list cost (AABB-same minus AAAA) was 236 ms, $F(1,18) = 204$, $p < .01$. The parceling of the alternation cost into these two components is partially

[6]The effect of key condition on the alternation cost was not due to the filtering out of trials in which the task and stimulus were both recently repeated since the difference in the alternation cost between key conditions is still 130 ms without this procedure applied.

validated by the lack of any difference between mean RT to stimuli from ABAB lists and AABB-different stimuli, $F<1$. These data do not support the hypothesis that the entire alternation cost occurs because the subject is in the wrong task set when each stimulus is presented. If this were the case then there should have been no mixed list cost at all. The fact that there is a shifting cost, however, is consistent with set switching and/or a "tuning" effect accounting for a large part of the alternation cost.

It will be recalled that the interaction of RSI with a factor tests whether the factor effects the RSI or baseline component. The mixed list cost was 255 ms at 0 ms RSI and 240 ms at 400 ms RSI, and this interaction was not significant, $F<1$. Thus, the mixed list cost would seem to be part of the baseline component of the alternation cost. In other words, there is no RSI component of the mixed list cost. RSI did have a significant effect on the shifting cost, $F(1,18) = 7.3$, $p < .05$, but the shifting cost did not vanish at the long RSI: the cost at 400 ms RSI was 158 ms, which was significant $F(1,18) = 62$, $p < .01$. These results indicate that the shifting cost has both a baseline and an RSI component.

Errors. The by-item error rates for Experiment 3 are shown in Table 5 as a function of key condition and sequence. Unlike previous experiments there was an overall effect of key condition, $F(1,18) = 5.2$, $p < .05$, task sequence, $F(2,36) = 8.7$, $p < .01$, and an interaction of the two variables, $F(2,36) = 4.2$, $p < .05$. Although the error rates are higher than in Experiment 1 and 2, they are still low by choice RT standards. In addition, the RT effects cannot be explained as speed accuracy tradeoffs since the slower conditions are the ones with the higher rates. Finally, AABB lists were not restarted on AABB-different stimuli significantly more or less often than on AABB-same stimuli, $F < 1$.

Summary of Experiment 3

Part of the alternation cost can be attributed to the subject having done a different task just before (the shifting cost). There is another part of the alternation cost, however, that appears to be due to the fact that two tasks are relevant on alternating lists compared to only one on pure lists (the mixed list cost). Both of these costs are large: 236 and 197 ms, respectively. At long RSI the mixed list cost is not reduced. Thus, more time does not allow subjects to overcome the difficulties of performing a task on a mixed list. The shifting cost, on the other hand, is reduced with RSI, but not eliminated. This means that the shifting cost can be divided into a baseline component and an RSI component, just like we divided the alternation cost into these components.

Response and Task Competition

We now turn to the hypothesis that the entire alternation cost is explained by some processing for the currently irrelevant task being done on the current stimulus (task competition). It is logically possible that the entire alternation cost reflects response or task competition. For example, subjects might be able to better suppress the inappropriate task after a longer RSI, accounting for the RSI component. It might further be that a task cannot be fully inhibited given even the long RSI's used in Experiment 1-3, but that actually doing task A, say, inhibits the set for task B more effectively, thus explaining the baseline shifting cost.

Experiment 1 provides a test of response competition models. In the Same-Keys condition of this experiment the color and form of the stimuli were sometimes associated with the same response key (e.g., a red "A" in Experiment 1) and were sometimes associated with different keys (e.g., a red "B" in Experiment 1). These stimuli will be called compatible and incompatible stimuli, analogous to compatible and incompatible stimuli in the Stroop task (Stroop, 1935). Since both tasks lead to the same response with compatible stimuli, response competition should not occur with them.

Figure 8A shows correct mean RT as a function of compatibility, alternation, and RSI in the same keys condition of Experiment 1. Compatibility had an overall effect of 65 ms, $F(1,17)=54$, $p < .01$. The compatibility effect was larger in the alternating condition (90 ms) than in the non-alternating condition (40 ms) and this difference was significant, $F(1,17)=6.5$, $p < .05$. The effect of compatibility was 89 ms at 0 ms RSI in the alternating condition and 91 ms at 1500 ms RSI in the alternating condition, $F < 1$.

Response competition is not reduced by the long RSI. Thus, response competition is not part of the RSI component. But this does not address whether response competition effects are in the mixed list cost or in the baseline component of the shifting cost. This is tested by Experiment 3. If response competition accounts for part of the shifting cost, then the effect of compatibility should be very much reduced on AABB lists in the case where the task repeats (AABB-same stimuli) compared to when the task alternates (AABB-different stimuli).

Figure 8B shows mean RT as a function of compatibility and RSI for AABB-same and AABB-different stimuli in Experiment 3. The overall effect of compatibility was 95 ms, and this was significant, $F(1,9) = 19$, $p < .01$, and as in Experiment 1 it did not interact with RSI, $F < 1$. Furthermore, the compatibility effect was not significantly effected by whether the previous task was the same or different from the current one, $F < 1$.

These analyses show that response competition effects are just as large on all stimuli on mixed lists, whether or not the task repeats or alternates, and whether or not a long or short RSI is provided. Thus, response competition effects account for part of the mixed list cost, but nothing else. However, response competition is not much larger in the alternating condition than in the non-alternating condition, so it does not contribute all that much even to the mixed list cost.

Experiment 4

Response competition is the most straightforward form of task competition. In fact, it is hard to concretely specify a task competition model that would not also predict that when the responses for the two tasks match, the "competition" will be gone. However, this does not mean that such a model is incorrect. In Experiment 4 we assess the role of task competition on alternating lists by producing a situation in which some of the stimuli are relevant to only one of the tasks. If these stimuli still show an alternation cost then it would indicate that there is more to this alternation cost than the application of the irrelevant task to the current stimulus.

In Experiment 4 the color and letter tasks from before were employed. A random 80% of the stimuli on each list were colored letters and the other 20% were white letters or colored discs. Thus, 80% of the stimuli were bivalent (relevant to both tasks) while the other 20% were univalent (relevant to only one of the tasks). Note that when a univalent stimulus occurred, it was always the stimulus called for by the sequence. If the alternation cost occurs because there are two conflicting sources of information on that stimulus, then the alternation cost should not be present for univalent stimuli. On the other hand, if all that is required is that before presentation the subject believes there might be two sources of information, then the alternation cost should still be found. So that we can compare RT to univalent stimuli to RT to compatible stimuli, only the Same-Keys condition was used. Finally, RSI was manipulated in order to determine whether any reduction in the alternation cost that might occur is due to a reduction of the baseline component, the RSI component, or both.

Methods

The methods used in this experiment were the same as in Experiment 2, except as noted.

Subjects. Twenty-four students at the University of California, San Diego participated for partial fulfillment of a course requirement.

Apparatus and Stimuli. The stimuli were colored letters, white letters, and colored circles. The white letters and colored disc will be called univalent stimuli, the colored letters will be called bivalent stimuli. The letters and colors were the same as in Experiment 3. The colored circle was the same as used in Experiment 2.

Design. Task Sequence varied between blocks of 5 trials. The order that the different task sequences occurred in the first 4 blocks was one of the 8 possible orders in which every other block was a non-alternating block and every other block was an alternating block. This ordering was then repeated 3 more times for a total of 16 blocks. The different block orderings were counter-balanced across subjects so that there were 3 subjects with each of the 8 possible block orderings.

Valence and RSI varied within each list. RSI was randomly chosen with equal likelihood from the two possible choices of 0 and 400 ms for each stimulus. There was an 80% chance for each stimulus to be bivalent, and a 20% chance to be univalent. Finally, key condition was not manipulated: all subjects were in the same keys condition.

Results and Discussion

Reaction Times. Mean correct RT in Experiment 4 is shown in Figure 9 as a function of RSI, alternation, and stimulus type (compatible, incompatible, or univalent). The overall alternation cost was 316 ms, $F(1,23) = 197$, $p < .01$. For bivalent stimuli the alternation cost was 369 ms, whereas for univalent stimuli the alternation cost was 263 ms. Although the alternation cost was significantly reduced in the univalent condition, $F(1,23) = 47$, $p < .01$, the 263 ms alternation cost for univalent stimuli was significant, $F(1,23) = 151$, $p < .01$. Thus, it would appear that the alternation cost still occurs even when task competition is no longer present.

In contrast to previous experiments, the effect of RSI on the alternation cost did not reach significance, $F(1,23) = 1.4$, $p > .2$. Note, however, that RSI still had an effect, $F(1,23) = 198$, $p < .01$, just not a significantly larger one on alternating lists. Although the overall effect of RSI on univalent stimuli was somewhat less than on bivalent stimuli (83 ms vs. 93 ms), this difference did not reach significant, $F<1$, and neither did the 3-way interaction between valence, alternation, and RSI, $F<1$.

The effects of response competition found in the present experiment are similar to previous findings: There was an overall 85 ms response competition effect, $F(1,23) = 91$, $p < .01$. Response competition was 128 on alternating lists and only 38 ms on non-alternating lists, and this difference was significant, $F(1,23) = 20$, $p < .01$. And finally, as before, RSI did not interact with the response competition effect, $F(1,23) = 2.3$, $p > .1$.

The extent to which task competition is a factor beyond response competition can be seen by comparing RT on univalent stimuli to RT on compatible stimuli. On non-alternating lists univalent stimuli were responded to a non-significant 10 ms slower than compatible stimuli, $F<1$, and this difference was not effected by RSI, $F<1$. Thus, univalent stimuli are a neutral condition to compatible and incompatible stimuli on non-alternating lists. On alternating lists, on the other hand, univalent stimuli were responded to an average of 48 ms faster than compatible stimuli, $F(1,23) = 11$, $p < .01$, and this difference was also not affected by RSI, $F < 1$. Thus, it appears that there is a small task competition effect beyond response competition. Since the task competition effect does not interact with RSI, it must be part of the baseline component of the alternation cost.

Errors. The by-item error rates were 3.4% and 2.2% for alternating and non-alternating task sequences, respectively (0.44 and 0.28 restarts per list). The effect of sequence was significant, $F(1,23) = 15$, $p < .01$. Thus, as in Experiment 3, subjects made more errors when alternating between tasks. But this cannot explain the alternation cost since more conservative responding on alternating lists should only make subjects slower on those lists and hence increase the effect.

Interim Summary

Experiments 3 and 4 increase our understanding of the alternation cost considerably. The result is that we can revise Table 4 proportionally. First, in Experiment 3 it was shown that the mixed list cost was additive with RSI whereas the shifting cost interacted with RSI. Thus, the entire RSI component of the alternation cost occurs precisely when the subject must do a different task from the stimulus before. In other words, the entire upper left corner of Table 4 has no cost associated with it. Furthermore, response competition was additive with RSI in Experiment 1 and 3 and was no larger on AABB-different stimuli than on AABB-same stimuli in Experiment 3. Therefore, response competition has its effects exclusively on the mixed list cost (upper right corner of Table 4). In Experiment 4 it was shown that task competition has only limited effects beyond response competition (48 ms), and these effects were limited to the baseline component.

Table 6 shows the revised mapping of the various possible reasons that the alternation cost occurs mapped onto the 2x2 division of the alternation cost as done in Table 4. Although not strictly ruled out by the data, task competition probably does not contribute to the baseline component of the shifting cost. This is because task competition is probably only present where response competition is. Thus it probably only contributes to the mixed list cost.

In Section One we argued that there were two components to the alternation cost -- the RSI and baseline components. This argument was partly based on the observation that it appeared as if alternating list performance would not be as fast as non-alternating performance even with a very long RSI. However, there was still a slight speedup between 800 and 1500 ms RSI in Experiment 2, so we cannot demonstrate conclusively that alternating performance would not eventually reach non-alternating list levels. But it was further argued that during the first 200 ms RSI alternating performance improved very much and thereafter much less. Thus, even if very long RSI resulted in equal performance on alternating and non-alternating lists, it would seem likely that there would be at least two underlying problems in alternating list performance -- one overcome within 200 ms and one much more slowly -- and these problems might be affected by different factors.

We now have additional justification for supposing that there are separate baseline and RSI components to the alternation cost. Figure 10 shows the pattern of interactions of RSI with different factors from Experiment 1, 3, and 4. If there were a single problem that occurs on alternating lists that was overcome quickly at first and thereafter more slowly, then the most plausible outcome is that RSI would interact with all factors that lead to partially overcoming this problem. However, three of the four factors are additive with RSI while the fourth, the shifting cost, interacts dramatically. Thus, a model in which at least two separate causes of slowing are independently affected by different factors is the most plausible.

The baseline component of the alternation cost.

It is interesting to try to account for the entire baseline component of the alternation cost. The pie chart in Figure 11 shows the alternation cost at 400 ms RSI in the same keys condition of Experiment 3. The entire alternation cost in this condition is 460 ms. The mixed list cost accounts for 302 ms of this. At 400 ms RSI there is a 158 ms shifting cost. Not all of this, however, is actually part of the baseline component. In Experiment 2 it was found that the alternation cost continues to decrease with RSI after 800 ms by 5 ms/100 ms RSI. Thus, about another 55 ms of the alternation cost at 400 ms RSI can be attributed to leftover RSI component. Thus, the baseline shifting cost (the "tuning" cost) accounts for about 100 ms of the alternation cost at this RSI. Finally, task and response competition account for 107 ms of the mixed list cost[7].

How might we account for the remaining 140 ms of the baseline component? First, it might be that there is slightly more of the RSI component reflected in the alternation cost at 400 ms than the 55 ms indicated above. This uncertainty arises because it is unclear whether in Experiment 2 RT had asymptoted as a function of RSI by 1500 ms RSI yet or not. Second, it might be that doing a task once does not get a subject fully ready to do that task again. Instead, more "tuning" might be needed to make processing as efficient as on non-alternating lists. Thus, the shifting component might actually be larger than the 100 ms arrived at above. Third, a criterion effect might account for some of the remaining 140 ms. Thus, because subjects are more

---

[7]The amount of the alternation cost that response and task competition accounts for was determined as follows: First, assume that if all other factors were controlled for that subjects would respond to univalent stimuli as fast on mixed lists as pure lists. Then from Experiment 4 we can estimate the cost and/or benefits of compatible and incompatible stimuli on both pure and mixed lists. So, subjects are 10 ms faster on pure lists when the stimulus is compatible than when it is univalent, and 32 ms slower when it is incompatible. Thus, since 4/7 stimuli are incompatible and 3/7 are compatible, subjects are 14 ms slower on non-alternating lists due to task and response competition. Since subjects are 48 ms and 176 ms slower on alternating lists when the stimulus is compatible and incompatible, respectively, rather than univalent, the cost of response and task competition is 121 ms on alternating lists. Thus, 107 ms of the shift cost is accounted for by response and task competition.

susceptible to errors -- maybe even because of task competition -- they select responses more conservatively. As a result, even when a univalent stimulus occurs (and task and response competition effects are eliminated), subjects are still slowed because they select responses more conservatively than on non-alternating lists.

The RSI component of the alternation cost.

The shifting cost interacted with RSI whereas the mixed list cost was additive with RSI. This suggests that the entire RSI component is a shifting cost. I.e., RSI has no benefit on mixed lists beyond the benefit on non-alternating lists, except when a shift of task set is required. But what occurs during the RSI when such a shift is required that allows subjects to respond faster?

One possibility is that some type of a discrete set switch occurs. As discussed above, this means that when the task changes, processing on the new task cannot proceed until the switch process is complete. At short RSI the time taken to implement this switch is included in RT whereas at long RSI the switch occurs during the RSI interval, hence the speed-up.

A switch of set need not be discrete in this way, however. What we have called a gradual set switch holds that as the RSI interval passes task set is in some way modified such that the switched-to task will be performed more efficiently, and this change of set occurs little by little. The gradual set switch is distinguished from the discrete set switch in supposing that once the stimulus is presented, processing on the task will begin. Thus, the speed-up in RT with longer RSI occurs because processing is more efficient at long RSI than at short RSI. A discrete switch, on the other hand, does not hold that processing efficiency varies at different RSI except in the sense that the subject is either in the right task set already or not.

It is interesting to ask how the set of the subject is modified by a set switch process. A set switch -- either discrete or gradual -- could correspond either to the activation of the switched-to task set or the inhibition of the switched-from task set, or both. In addition, it might also be accompanied by a criterion change due to less susceptibility to errors as the switched-to task set is activated and/or the switched-from task set is inhibited.

A set switch supposes that somehow the machinery that selects responses becomes more efficient at selecting responses for the switched-to task during the RSI. It might be, however, that nothing of this sort occurs. Instead, perhaps all that happens is that the subject decides which task the machinery will be used for. In the non-alternating condition, however, no such decision needs to be made since the same task is done throughout a list. This is the set decision hypothesis considered above.

One or more of these hypothetical processes might occur during the RSI on alternating lists. The data up to this point do not help to discriminate between the various possibilities. In Experiments 5-7 we further explore task switching in order to get a better handle on what occurs during the RSI.

Experiment 5

In Experiment 5 we further characterize the RSI component of the alternation cost. Subjects made button press responses to the color and forms of colored letters. As before, they either alternated between the two tasks or performed the same task on each item in a list. Somewhere between the 5th and 10th item, however, the stimulus was univalent. Half the time it was for the task called for by the preceding task sequence and half the time it was not, and subjects performed whichever task was relevant. Thus, the task sequence might be color-letter-color-letter followed either by a white letter or by a colored circle. In addition, RSI was 0 ms in some blocks and 400 ms in others. There are two key issues here: 1) Will subjects be aided when instead of having to do the alternate task as they expect they have to do the same task they just did? 2) What will the effect of RSI be when it is the wrong task that they are preparing for?

A gradual set switch model supposes that the subject is better prepared for the switched-from task at the beginning of the RSI, but better prepared for the switched-to task more and more towards the end of the RSI. It therefore naturally explains how subjects might be better off getting an unexpected same task than an expected different task at 0 ms RSI, but the reverse be true at the end of the RSI. However, it is also possible that a subject cannot immediately begin selecting a response for a task that was not expected. It might be that instead there is a delay before the appropriate task can be applied. Even in this case, however, more RSI ought to hurt the subject on an unexpected same task since the RSI interval is spent preparing for the wrong task.

A discrete set switch or a set decision model, on the other hand, makes different predictions. Both of these models assume that the RSI is used to commit the subject to one of the two tasks. Thus, if we assume that this operation cannot be terminated prematurely, then getting a stimulus for an unexpected task should slow the subjects down considerably. In fact, if an unexpected task occurs subjects would have to commit first to the expected task and then to the unexpected but now correct task. Because of this, these models predict that RSI will help the subject respond faster. This is because the first of the two switches/decisions can occur during the RSI.

Thus, a gradual set switch, a discrete set switch, and a set decision model are all consistent with the finding that when the univalent stimulus unexpectedly repeats the previous task, subjects are slower than when it is the (expected) alternate task. In addition, a gradual set switch is also consistent with the opposite outcome. Finally, whereas the other two possibilities predict that RSI will not hurt subjects performance when the univalent stimulus is for the unexpected task on alternating lists, a gradual set switch predicts that it will.

Methods

The methods used in this experiment were the same as in Experiment 3, except as noted.

Subjects. Thirty-two undergraduate students at the University of California, San Diego participated for partial fulfillment of a course requirement.

Stimulus Lists. Each list consisted of the presentation of five to ten stimuli. The last stimulus in each list was univalent, all the others were bivalent. Otherwise, the lists were generated just as in Experiment 3.

Design. Task Sequence and RSI varied between blocks of ten trials each. There were four Task Sequences (color, letter, color-letter, and letter-color) and two RSI's (0 and 400 ms). The task sequence of the first four blocks was chosen from one of the eight possible permutations of the four task sequences such that every other block had alternating lists and every other block had non-alternating lists. This block ordering was then repeated three more times for a total of 16 blocks. RSI was randomly determined for the first four blocks, the complements were used for the next four, the values used in blocks 5-8 for the next four, and the values in blocks 1-4 for the final four blocks. One item in each list was univalent. This item was always the last item, and was randomly selected from position five to ten. Finally, the order in which the task sequences occurred across blocks and key condition was counter-balanced across subjects.

Results and Discussion

Reaction Times. Figure 12 shows correct RT to bivalent stimuli as a function of alternation and RSI in Experiment 5. Key condition did not interact with any variable, so it is averaged across in the figure and not reported in the analyses that follow. The data from bivalent stimuli paralleled previous findings: There was an overall alternation cost of 328 ms, $F(1,30) = 247$, $p < .01$, and an overall RSI benefit of 104 ms, $F(1,30) = 120$, $p < .01$. The alternation cost at 0 ms RSI was 380 ms and the alternation cost at 400 ms RSI was 284 ms, and this difference was significant, $F(1,30) = 16$, $p < .01$.

When a univalent stimulus occurred in this experiment, subjects no longer performed the task determined by the instructed task sequence, but instead performed whichever of the two tasks was relevant to the univalent stimulus. Figure 13 shows correct RT to univalent stimuli in the 0 ms RSI condition as a functions of bivalent sequence and whether the univalent task was the same as the previous task or not. Subjects were overall 132 ms faster on non-alternating lists than on alternating lists, $F(1,30) = 87$, $p < .01$. When the univalent task was the same as the previous (bivalent) task, RT was an average of 102 ms faster than when it was different, $F(1,30) = 38$, $p < .01$. This effect, however, depended on whether the bivalent part of the list was alternating or not: when it was non-alternating, there was a much larger 372 ms effect, and when it was alternating there was a 148 ms effect in the opposite direction. This interaction between alternation and same vs. different task was significant, $F(1,30) = 87$, $p < .01$. One way to look at these findings is that subjects were faster when the univalent stimulus was for the task that continued the bivalent task sequence. In particular, although in Experiment 3 it was found that there was an advantage of doing the same task two times in a row, there was no overall advantage here. In fact, the slowest overall condition was when the bivalent sequence was alternating and the univalent task was the same as the previous task.

As noted previously, a gradual set switch is consistent with this result if one assumes that there is some delay that occurs when the stimulus is for an unexpected task. However, this model also predicts that as RSI increases, subjects should become even slower when an unexpected repeated task occurs. Figure 14 shows the RSI benefit found in the four univalent conditions. RSI provided a benefit of 54 ms when the univalent task continued the bivalent sequence compared to a 50 ms benefit when it did not, $F<1$. The interaction of this variable with whether or not the bivalent sequence was alternating was not significant, $F(1,30) = 1.1$, $p > .3$ nor was the overall effect of bivalent sequence on the RSI benefit, $F<1$. In short, lengthening RSI always helps and approximately to the same extent, even when subjects are expecting the wrong task at the end of it. This strongly argues against a gradual set switch.

Errors. The mean number of times each list had to be re-started because of an error is shown in Table 7 as a function of task sequence, RSI, and key condition. By-item error could not be computed because lists varied in length. Key condition and task sequence were both significant, $F(1,30) = 4.5$, $p < .05$, $F(1,30) = 19$, $p < .01$, respectively, as was also true in Experiment 4. The interaction of these variables was not significant, $F(1,30) = 1.5$, $p > .2$. RSI did not have an effect on the number of restarts, $F<1$, although there was a significant interaction between RSI and key condition, $F(1,30) = 4.2$, $p < .05$. There were no other significant interactions. Although, as in previous experiments, there do seem to be significant effects on error rates, conditions with higher rates in general correspond to conditions with slower RT, thus speed-accuracy tradeoffs are not a concern.

One question unanswered by the above is whether subjects are more or less likely to make an error on the univalent stimuli. If errors are just as likely on univalent as bivalent stimuli then 14.1% of the trial restarts should occur on univalent stimuli. On non-alternating lists 22.1% and 14.7% of the restarts were on univalent stimuli, for

0 and 400 ms RSI's, respectively. The first of these two figures was significantly different than 14.1%, $F(1,31) = 5.4$, $p < .05$, but the second was not, $F<1$. On alternating lists 11.6% and 15.2% of the restarts were on univalent stimuli for 0 and 400 ms RSI's, respectively. Neither of these were significantly different from 14.1%, $F=1.0$ and $F<1$. Thus, although these figures do not give a clear answer one way or the other, they do suggest that at least on alternating lists subjects are not much more or less likely to make an error on a univalent stimulus.

Hybrid Models

One aspect of the data that should not be overlooked is that on alternating lists there was a smaller RSI benefit when responding to univalent stimuli than when responding to bivalent stimuli (54 ms vs 116 ms). A discrete set switch or a set decision model would most naturally predict that the RSI benefit would be the same on univalent and bivalent stimuli. This is because these models hold that the benefit is found for the same reason for both univalent and bivalent stimuli: the decision to respond to the expected task or the switch to this task occurs during the RSI.

Is this pattern of results consistent with a hybrid model, say a discrete set switch in parallel with a gradual switch, each switch accomplishing a different aspect of preparing for a task? A hybrid model of this sort would explain a smaller benefit on alternating lists when the univalent task does not continue the sequence. That is, the RSI effect would be the sum of the benefit derived from performing the discrete switch during the RSI and the cost of gradually preparing for the wrong task. However, a hybrid model would still predict a large RSI benefit when the univalent task continues the alternating sequence, and this did not occur. It would seem, then, that (hybrid models true or not) an explanation of the small RSI benefit must be sought elsewhere.

One possibility is that subjects sometimes recognize that the stimulus is univalent before they recognize which task it is for, and at this point re-set themselves from "scratch", i.e., even if it means performing a switch to the set they are already in. On such trials, there would be no RSI benefit since the setting for the appropriate task begins when the stimulus is presented no matter what the RSI.

Nonetheless, the main point should not be overlooked: when the RSI can only be spent preparing for the wrong task, RSI resulted in a benefit and not a cost. This is at odds with a gradual switch. In addition, subjects were slower on alternating lists when the univalent stimulus did not continue the alternating sequence than when it did. This is predicted by models that suppose the RSI component of the alternation cost corresponds to some process that commits the subject to one of the two tasks, and when the subject commits to the wrong task a switch back is necessary. A discrete set switch and the set decision model are examples of this. This is not to say that a hybrid model is completely ruled out. But, such a model offers no additional explanatory power over a discrete set switch or a set decision model. Moreover, the gradual set switch in a hybrid model would seem to play only a small role since the RSI benefit is only slightly smaller on unexpected repetitions (alternating lists, previous task same) than on expected alternations (alternating lists, previous task different). Thus, we can conclude that if there is a gradual set switch it is not the only or even main underlying cause behind the RSI component.

Experiment 6

In the Introduction it was noted that Spector and Biederman (1976) found that when subjects alternate between disjoint tasks -- tasks that use disjoint stimulus sets -- there is little or no alternation cost. This presents a bit of a puzzle: In Experiment 3 we found that set shifting has sizeable effects on both the baseline and RSI components of the alternation cost, and this factor (unlike response and task competition, for example) would not seem to go away just because the stimulus sets are disjoint. As mentioned in the Introduction, one solution to this puzzle would be if when the stimulus sets are disjoint, subjects can form a task set in which they are ready for both tasks at once. When all the stimuli in a list are bivalent, however, the task set must exclude the irrelevant task (or else the subject would be extremely error prone), thus requiring a set switch each stimulus.

In Experiment 6 we test a case that is intermediate between disjoint tasks and tasks with bivalent stimuli. The first six stimuli in each list were bivalent; the next four were univalent. Subjects either alternated between the letter and color task on the first six stimuli or performed one task on all stimuli. Whether each univalent stimulus was a letter or a color was randomly determined for each stimulus independently, and the subject performed whichever task was relevant on these stimuli. Thus, it might be that on the univalent part of the list subjects will no longer have to be ready for just one task at a time, thus avoiding a set switch when the task changes. On the other hand, it might turn out that subjects will still incur a switching cost .

The outcome of this experiment will help us better understand what occurs during the RSI. For example, suppose that the RSI is spent inhibiting the switched-from task to undo the "tuning" effect. That is, doing a task puts the subject in a state in which they are ready for that task again (the "tuning" effect) and the RSI on alternating lists is spent inhibiting the task set for the task just done so that it does not interfere with the selection of the next response. Suppose further that the subject can be ready for both tasks so long as each has been performed since inhibition was turned off. Then on the univalent part of the list , where this inhibition can be turned off, there should be no more cost associated with changing tasks after each task has been performed once.

Methods

The methods used in this experiment were the same as in Experiment 5, except as noted.

Subjects. Thirty-two undergraduate students at the University of California, San Diego participated for partial fulfillment of a course requirement.

Stimulus Lists. The lists were generated just as in Experiment 5, except that list positions 1 through 6 were always bivalent and positions 7 through 10 were always univalent. Which task each univalent stimulus was for was randomly determined.

Design. The design was identical to Experiment 5, except that RSI was fixed at 0 ms throughout and there were only 8 trials in each block.

Results and Discussion

Reaction Times. The data from the bivalent part of the lists replicated previous findings. RT was 803 ms and 1214 ms on non-alternating and alternating lists, respectively, and this 409 ms alternation cost was significant, $F(1,30) = 297$, $p < .01$. As in Experiment 3 there was a significant effect of key condition on the alternation cost, with a 465 ms alternation cost in the same keys condition and a 356 ms alternation cost in the different keys condition, $F(1,30) = 5.3$, $p < .05$.

Figure 15A shows mean correct RT on univalent stimuli as a function of list position (7-10) and bivalent sequence (alternating or non-alternating) when the last bivalent task plus the univalent sequence up to that point was non-alternating. The two dotted lines in the figure refer to mean RT to bivalent stimuli for both alternating and non-alternating bivalent sequences. When the bivalent sequence is alternating and the first univalent stimulus does not continue the sequence subjects are 161 ms slower than they are on the bivalent part of the list. This is similar to what was found in Experiment 5. There was a significant effect of the bivalent sequence, $F(1,30) = 40$, $p < .01$, but this is due to the difference on response number 7, with an average difference of only 5 ms for the other 3 stimuli; the interaction of bivalent sequence and list position was significant, $F(1,30) = 50$, $p < .01$. In addition, the overall effect of list position was significant, $F(3,90) = 97$, $p < .01$. There was also a significant interaction of bivalent sequence and key condition, $F(1,30) = 7.8$, $p < .01$, paralleling that found on the bivalent part of the lists.

Figure 15B shows mean correct RT on univalent stimuli as a function of list position and bivalent sequence when the last bivalent task plus the univalent sequence was alternating. The data from one subject is not included in the figure or the analyses below because this subject had no cases in which the univalent sequence happened to be alternating all the way to the final stimulus. RT dropped from 1239 ms on the first univalent stimulus to 926 ms on the last, or from 23 ms slower than bivalent alternating RT and 433 ms slower than bivalent non-alternating RT to 283 ms faster than bivalent alternating RT and 113 ms slower than bivalent non-alternating RT. The effect of list position was significant, $F(3,87) = 50$, $p < .01$. Neither the effect of bivalent sequence nor its interaction with list position were significant, both F's < 1.

These data indicate that subjects do not immediately go into a "combined" task set in which they can respond to both color and letter stimuli without an alternation cost. That is, even in list position 10 there is an alternation cost. The data in Figure 15C back this up further. This figure shows mean correct RT on univalent stimuli as a function of response number, bivalent sequence, and whether the previous task was the same or different, regardless of the univalent sequence before that. It will be noticed that the difference between same versus different previous task is larger in Figure 15C than in Figure 15B. The reason for this will become more clear below.

Errors. Table 8 shows the by-item error rates for alternating and non-alternating lists in Experiment 6 broken down by valence[8]. The error rate to bivalent stimuli was 1.0% higher than the error rate to univalent stimuli, and this difference was significant, $F(1,30) = 18$, $p < .01$. In addition, the error rate on alternating lists was higher than the error rate on non-alternating lists by 0.9%, and this difference was also significant, $F(1,30) = 32$, $p < .01$. These two variables significantly interacted, $F(1,30) = 17$, $p < .01$, corresponding to the fact that the majority of the valence effect was on alternating lists. Finally, key condition was not significant, $F < 1$, but did significantly interact with valence, $F(1,30) = 5.3$, $p < .05$, indicating that the majority of the valence effect occurs in the same keys condition. No other interactions with key condition were significant. As in other experiments, higher error rates corresponded to slower RT's, so speed-accuracy tradeoffs are not an issue.

The inhibition model

The inhibition model discussed above makes very specific predictions here: On the univalent part of the list there is no need for inhibition, so once each task has been performed on this part of the list both sets should be loaded and the alternation cost should go away. However, as is apparent from Figure 15, there is a sizable alternation cost even on the last univalent stimulus. So it does not seem to be the case that subjects can simply turn off some inhibition of inappropriate mappings and thereby keep both mappings fully activated.

---

[8]The per item error rate for univalent stimuli is computed with the formula:

uni-error = 1 - 4th-root of 1/(1+uni-restarts).

For bivalent stimuli, a correction is first applied to the restarts:

bi-restarts' = bi-restarts * (1-uni-error).

The per item error rate for bivalent stimuli is then:

bi-error = 1 - 6th-root of 1/(1+bi-restarts').

See Appendix for details.

One might think, though, that it worked in a slightly different way. It might be that each time a subject does a task on the univalent part of the list, that task is loaded a bit more. Eventually after doing it several times it becomes "fully" loaded. Moreover, it might be that preparing for one task makes the subject less prepared for the other, i.e., preparation is a zero sum game[9]. To test this possibility we looked at RT on the final univalent stimulus (list position 10) as a function of the preceding sequence. If the task 1-back and 2-back (list position 9 and 8, respectively) were the same as the present one subjects were on average 175 ms and 162 ms faster, respectively, than if the corresponding tasks were different. These effects were significant, $F(1,30) = 121$, $p < .01$, and $F(1,30) = 116$, $p < .01$, respectively, and there was no significant interaction, $F<1$. These effects are consistent with the above hypothesis. However, if the task 3-back (list position 7) was the same subjects were an average of 41 ms slower, $F(1,30) = 15$, $p < .01$. The two-way interactions with the other two variables were not significant, but the 3-way interaction was, $F(1,30) = 22$, $p < .01$. It would seem hard to reconcile the harmful effect of having done the same task 3 stimuli back with a bit by bit loading of task set.

Set expectation effect

Why is there a harmful effect of having done the same task 3 stimuli back? It turns out that the entire pattern of effects is well explained as a set expectation effect. That is, subjects expect one task or the other and pay a cost when they are wrong. To make this point clear we will first consider some data from 2-choice serial RT tasks. In a 2-choice serial RT task subjects are presented with one of two stimuli (say A and B) on each trial, and make one button press response if they see an A and a different

---

[9]One might wonder if preparation were really a zero sum game why there would be an alternation cost with disjoint tasks. However, this situation might only arise when some of the stimuli are bivalent.

one if they see a B. The next stimulus is presented some time after the subject makes a response to the previous stimulus. Figure 16A shows data from a 2-choice serial RT task with short RSI (50 ms from key up) and long RSI (50 ms from key down) as a function of the previous stimulus sequence from Vervaeck and Boer (1980, Experiments 1 and 2, respectively). These will be called stimulus expectancy curves. The sequences are laid out in the graph so that from left to right the recency and number of alternations in the sequence increases. For short RSI, RT increases from left to right. Thus, the more the recent sequence involves repeated elements, the faster RT to the next stimulus will be. Even if the sequence up to this point has been alternating, subjects are faster when the next stimulus is a repetition. However, with longer RSI's, during which subjects are presumably able to build up expectations for which stimulus will occur, RT first increases and then decreases from left to right. In short, the more the sequence resembles either alternating or non-alternating the more one is helped if the next stimulus continues this sequence and the more one is hurt if it does not.

The same general pattern of effects found at long RSI in 2-choice serial RT is found in Experiment 6, except that anticipation of the next task takes the place of anticipation of the next stimulus. Figure 16B shows RT to the last univalent stimulus on each list as a function of the univalent task sequence up to and including this stimulus (set expectancy curves). These data strongly resemble the long RSI data in Figure 16A. Thus, just as Vervaeck and Boer's subjects appear to expect one stimulus or the other when given a long RSI, our subjects expect one task or the other, and apparently do something to prepare for that one at the expense of the other.

It seems logical to suppose that the process that commits subjects to one task or the other over the RSI on alternating lists is the same one that is applied to commit subjects to one task or the other depending on the preceding sequence here. If this is the case, however, it is difficult to understand why RT_{ABAB} is so much faster than RT_{AAAB}. Both entail a switch from the task just performed to the alternating task. One might suppose that the difference arises because the switch in the AAAB case is not performed until it is recognized that the stimulus is not for task A as the subject had gambled, but in the ABAB case the switch is started right after responding to the previous task, i.e., the switch in the ABAB case gets a head start. However, the difference is over 200 ms; it seems unlikely that subjects simply do nothing for 200 ms when the preceding three tasks are different from the present one.

There are other problems for this account too. When the sequence is ABAB there should be both a baseline shifting cost and an RSI shifting cost. However, RT_{ABAB} - RT_{AAAA} is only 118 ms -- much smaller than the sum of the baseline and RSI components of the shifting cost on alternating lists. One might explain this by supposing that there is no baseline component on the univalent part of the list, but it is unclear why this would be so.

There are still more problems. Consider the ABAB and ABAA sequences. These correspond to the same task sequence on response number 7-9 and only differ in the task that occurs on response number 10. Presumably, the fact that RT_{ABAB} is less than RT_{ABAA} indicates that subjects tend to switch to the alternate task set when the preceding sequence alternates. Similarly, that RT_{AAAA} is faster than RT_{AAAB} indicates that when the preceding sequence repeats subjects tend to remain set for that task. Thus, when the sequence is ABA* (where the asterisks could be either A or B), subjects tend to make an extra switch compared to AAA*, with no advantage on average. Thus, RT_{AAA*} should be much faster than RT_{ABA*}. Yet, RT_{ABA*} is slightly faster: 979 ms versus 996 ms, $F < 1$.

All of this makes sense, however, if we suppose that subjects are able to set themselves for the task expected based on the sequence without a time cost. Thus, after ABA subjects suffer only the baseline component if task B occurs and only the RSI component if task A occurs. After AAA subjects suffer both components if task B occurs and neither if task A occurs. This hypothesis makes an additional prediction: RT_{AAAB}-RT_{ABAA} = RT_{ABAB} - RT_{AAAA}, where the quantities on either side of the equal sign correspond to the baseline component of the shifting cost. The numbers on the left and right of the equal sign are 146 ms and 131 ms, respectively, and they are not significantly different, $F<1$. We do not, however, have much power in this comparison since each of the four means above is based on RT to one stimulus every eight lists (standard error of the mean for the comparison between the four means -- the difference between the quantities on either side of the equal sign -- is 35 ms).

So what is it that occurs during the RSI on alternating lists? These results suggest that it is some process that does not necessarily take any time, but does on alternating lists and in the case, above, where the task was not the one expected based on the preceding task sequence. This is not congenial with a set switch. Why would a set switch take no time in some situations? For example, if the switch were analogous to loading a set of rules for the upcoming task or changing the weights in a neural network that selects the response for the task, it does not make sense that it usually takes a couple hundred milliseconds but under other conditions takes no time at all.

The set decision model, however, fits in quite well with this data. The critical idea is that the subject must be set for one task or another, but that this setting does not take long to accomplish. What does usually take time, however, is the decision of what task to be set for. On alternating lists this amounts to a memory retrieval for what task is next. However, when the sequence is random and the stimuli are univalent, the initial decision depends on the preceding task sequence and is accomplished about as quickly no matter what the sequence. If this decision is wrong -- which is the case half the time -- a new decision must be made to get set for the other task, and the duration of this decision includes the time taken to "notice" that the stimulus is for the other task. There is, however, still a baseline cost of alternating, which explains why the right half of Figure 16B is raised by about 120 ms over the left half (in particular, ABAB over AAAA).

In short, the set expectancy curves are well explained if we assume that the RSI component of the shifting cost corresponds to a process that commits the subject to one task or another and that this process can be completed without large time costs when it is determined by the preceding sequence of stimuli. This also explains why the alteration cost shown in Figure 15B was much smaller than the effect of doing the same versus different task on the previous stimulus in Figure 15C -- the alternation cost in Figure 15B only reflects the baseline component of the alternation cost because subjects anticipated that the sequence would alternate. A set switch is inconsistent with these findings because it should take just as long to load a new set when the stimuli are univalent as it does when the stimuli are bivalent. The task set decision model, however, is consistent with this, since by this hypothesis it does not take a large amount of time to commit to one set or another, but it is the decision of which task to commit to that takes time, and there is no reason that this decision cannot sometimes be very quick.

Experiment 7

In Experiment 3 we found evidence that doing two of the same task in a row results in better performance on the second one. In Experiments 5 and 6 this did not occur, but presumably this was because subjects were expecting the other task and began performing some type of switch which they could not halt. One might think that in Experiment 6 subjects should have known not to perform this switch since the first univalent stimulus was always the 7th in the list. However, subjects may not have kept track of where in the list they were. In Experiment 7 as in Experiments 5 and 6 subjects alternated between two tasks until one stimulus occurred that was univalent for one or the other of the two tasks. But in this experiment, everything was done to assure that the subject would know which stimulus would be univalent, so that the subject would not commit to one task or the other. First, there were always five stimuli in a list with the last being the univalent one. Second, the first four stimuli occurred around the circumference of a circle in the clock-wise direction, always starting at the top of the circle. The last stimulus (the univalent one) always occurred in the center of the circle. Thus, if subjects are able to get into a state in which they are not committed to either task, then this experiment would seem to provide those conditions.

Methods

The methods used in this experiment were the same as in Experiment 5 except as noted.

Subjects. Sixteen undergraduate students at the University of California, San Diego participated for partial fulfillment of a course requirement.

Stimulus Lists. Lists consisted of 5 items on each trial, the first four colored letters (bivalent) and the last one either a colored disc or a white letter (univalent).

Procedure. In previous experiments stimuli were presented on the screen from left to right. In Experiment 7 the first 4 stimuli were presented 4.5 cm above, to the left of, below, and to the right of, respectively, the center of the screen. The fifth stimulus was presented in the center of the screen.

Results and Discussion

Reaction Times. Figure 17 shows correct RT to bivalent and univalent stimuli as a function of whether the bivalent part of the list was alternating or non-alternating. Univalent stimuli are further broken down by whether the univalent task was the same as or different from the previous task (univalent-same and univalent-different, respectively). Key condition once again did not significantly interact with any other variables and so is averaged across in the figure and not reported in the analyses below. On the bivalent part of the list subjects showed an alternation cost of 383 ms, raising RT from 746 ms in the non-alternating condition to 1129 ms in the alternating condition, and this difference was significant, $F(1,14) = 283$, $p < .01$.

On the surface, the pattern of results for the univalent stimuli resemble what was found in Experiments 5 and 6. When the univalent stimulus was from an alternating list subjects were on average 60 ms slower than when the stimulus was from a non-alternating list, $F(1,14)=12$, $p < .01$. When the stimulus was for the same task as had just been performed subjects were on average 89 ms faster, $F(1,14) =26$, $p < .01$, but as in Experiment 5 this variable interacted with list type, $F(1,14)=125$, $p < .01$: subjects were 235 ms faster doing the same task twice in a row when the bivalent sequence was non-alternating, but 57 ms slower when the bivalent sequence was alternating. Thus, as before, subjects are better off when the univalent stimulus continues the sequence, even though they are told that the stimulus is no more likely to continue the sequence than not.

Despite the similarities, there are some crucial differences between the findings of Experiment 7 and the previous two experiments. First, in Experiment 5 and 6 subjects were much worse on univalent-same stimuli from alternating lists than they were in any other condition. In Experiment 7, however, subjects were faster in this condition than they were to respond to the bivalent stimuli on alternating lists and only

57 ms slower than to respond to univalent stimuli that continued the alternating sequence. Thus, although in this experiment it is still better to get a univalent stimulus that continues the task sequence, the cost is not nearly as large in this experiment when it does not as in Experiments 5 and 6. In fact, it is better to get a univalent stimulus that does not continue the sequence than to get a bivalent stimulus that does. Below we will explore reasons for why the differences between Experiments 5-7 may exist.

Errors. The by-item error rates in Experiment 7 are shown in Table 9 as a function of task sequence and valence. Subjects made significantly more errors on univalent stimuli than on bivalent stimuli, $F(1,14) = 10$, $p < .01$, 4.4% vs. 2.9%. This difference was significantly larger on non-alternating lists than on alternating lists, $F(1,14) = 5.5$, $p < .05$. There were no other significant main effects or interactions on the by-item error rates.

Ready, Set, Go! model

At this point we are able to put forth a reasonable working hypothesis that can explain our findings. This model will be called the Ready, Set, Go! model (RSG model). The model is described by the following 6 properties:

RSG1. There is a mechanism that selects responses, called the response selection mechanism, and this is the only mechanism that can select responses.

RSG2. This mechanism can be in a state that ranges from being not ready to do a particular task (and possibly ready to do a different one) to being fully ready for that task.

RSG3. The readiness for a particular task cannot be changed during "free" time (the RSI).

RSG4. The response selection mechanism is made ready for a task by that task being the last task performed (readiness may slowly degrade with time, however).

RSG5. Before the response selection mechanism is used for a task, a decision as to which task it is to do must be made. This decision of which task to do -- setting the mechanism -- is not the same as being ready for a task. The response selection mechanism can be set for one task and ready (in the sense of RSG2-RSG4 above) for a different one.

RSG6. The setting of the response selection mechanism for one task or another is done during the RSI, and cannot be interrupted once started.

Note that the use of the term "readiness" in RSG2-RSG4 above is completely analogous to the use of the term in Section One. The only difference is that here it is applied only to the response selection mechanism whereas in Section One it is applied to the whole subject.

In summary, there are two components to being prepared for a task: readiness of the response selection mechanism for the task, and the setting of this same mechanism for the task. These correspond to a "tuning" effect and a set decision, respectively. The mechanism must be set before task processing begins, but does not necessarily have to be ready beforehand (although processing will be faster if it is). Setting the mechanism is achieved during the RSI, and is completed -- on average -- within a few hundred milliseconds. Readiness cannot be changed during the RSI, therefore the readiness of the response selection mechanism is reflected in the baseline component of the alternation cost. However, this is not the whole baseline component. For example, it was shown that there is more response and task competition on alternating lists than on non-alternating lists, and this difference does not disappear with large RSI's. In addition, there might be a criterion effect at play here, accounting for some of the baseline component.

Let us consider the findings that this model explains and that must be explained by any viable model:

1. The alternation cost has two components: the RSI component which is overcome during an RSI, and the baseline component which is not (Experiment 1 and 2). This is explained by the RSG model since setting of the mechanism occurs during the RSI but readying of the mechanism does not. In addition, several other factors contribute to the baseline component of the alternation cost.

2. Repeating a task twice in a row helps a lot even when plenty of "free" time is given to get ready for the upcoming task, i.e. even at long RSI there is a considerable benefit of responding to AABB-same stimuli over AABB-different stimuli (Experiment 3). This is explained by RSG4.

3. All of the additional benefit of RSI found on alternating lists over non-alternating lists is a shifting cost and not a mixed list cost (Experiment 3). The exact explanation of this depends on the reason that an RSI benefit is found on non-alternating lists. If one supposes that the entire RSI benefit on non-alternating lists is due to a "partial" refractory period as suggested by Wilkinson (1990), then the extra RSI benefit on

alternating lists is due to the fact that the set decision takes longer than this "partial" refractory period. In addition, if one further assumes that no set decision is required on AABB-same stimuli, then this also explains why the benefit is only as large as on non-alternating lists in this case.[10]

If one instead supposes that there is a full refractory period following a response during which nothing except perhaps recognition of the stimulus is accomplished, then the RSI benefit on alternating lists is the sum of this delay and the delay associated with deciding which task to do next. As before, the relatively small AABB-same RSI benefit is explained if no set decision is required on these stimuli.

One might also reverse the logic and assume that whatever occurs on the alternating lists during the RSI also occur on non-alternating lists during the RSI, just to a lesser extent. Thus, it might be the case that a set decision must be made not only on alternating lists but on non-alternating lists too. (This account would be similar to an "unprepared" period). In this case the difference in RSI benefits would arise because it is much quicker to decide to do the same task each time (a non-decision) than to choose between two different ones. Thus, the larger RSI benefit on alternating lists would occur because the set decision involves one more bit of information (Hick, 1952; Hyman, 1953).

4. <u>When a univalent stimulus is presented at the end of a list of bivalent stimuli, and the univalent stimulus sometimes does and sometimes does not continue the bivalent sequence, subjects are always faster in the former case (Experiments 5-7)</u>.

The RSG model explains this by assuming that subjects perform a set switch on the univalent stimulus before trying to select a response. Thus, if the stimulus is

---

[10]If it does not seem reasonable that one does not have to make a set decision on AABB-same stimuli, try saying "AABBAABBAABB..." as fast as possible yet at an even pace

actually for the same task as just done they must switch back. One might think that the RSG model would predict one of two other possibilities: 1) Subjects switch to a set from which both tasks can be done (as one might assume is done when the tasks are disjoint). However, it is not clear how subjects form task sets, and they might not have such a set available. We will have more to say on this in Section Three. Or, 2) Subjects stay in whatever set they are in, and switch if the stimulus is for the other task. This would, after all, entail an average of one less switch than if they automatically switch into the alternate set first. A pilot experiment based on Experiment 7 is relevant here. In this experiment four bivalent stimuli were presented around a circle followed by a single univalent stimulus in the center of the circle, as in Experiment 7. The difference was that in the pilot experiment the univalent stimulus was always for the same task. So the subject would know beforehand that the sequence would be color--letter--color--letter--letter, for example. Nonetheless, in this experiment the subjective feeling was that one was trying to continue the sequence. One simply forgets to try to do two letter tasks in a row by the end of the sequence. Thus, we do not find it unintuitive that subjects make what is really, on average, one extra switch beyond what is needed.

One thing of interest here is that the more of a "surprise" the univalent stimulus is, the slower subjects are when it is for the unexpected task. In Experiment 5 the univalent stimulus occurred sometime between the 5th and 10th stimulus. Thus, subjects could not be certain when it would occur unless it had not occurred on the first nine stimuli. In this experiment, when the bivalent sequence was alternating but the univalent stimulus did not continue this sequence, subjects were 175 ms slower than when alternating on the bivalent part of the list. In Experiment 6 the first univalent stimulus always occurred on the 7th stimulus. However, if subjects did not keep explicit track of where they were in the list they would not know the univalent stimulus was about to occur. In this case subjects were 161 ms slower when the univalent stimulus did not continue a bivalent alternating sequence than they were on the bivalent part of an alternating list. In Experiment 7 where the bivalent stimuli appeared around the circumference of a circle and the univalent always occurred in the center, subjects were 51 ms faster on a univalent stimulus that did not continue an alternating sequence than they were on the bivalent stimuli in the same list (although the subject was still faster if the univalent stimulus continued the sequence).

The RSG model explains these findings straightforwardly. The response selection mechanism must be set for a task before a response can be selected. When a univalent stimulus occurs that does not continue the sequence, the mechanism must be re-set for the task that the stimulus is for. This re-setting itself does not take a large amount of time, nor does setting the mechanism take a large amount of time when alternating between tasks on a bivalent list. What does take time is the decision to set the mechanism in a particular way. When alternating on a bivalent list this involves a retrieval from memory of what the next task is. When one is set for the wrong task after a univalent stimulus has been presented, the decision amounts to noticing that the stimulus is for the other task. And this decision is very affected by how predictable it was that the current stimulus was going to be univalent.

5. When univalent stimuli are presented at the end of a sequence of bivalent stimuli as in 4, more RSI helps, even when the additional time is applied to a switch toward the wrong task (Experiment 5). This is a simple consequence of the RSG explanation of 4: Since the setting the response selection mechanism to the alternate task is performed in all cases, and can occur during an RSI, RSI helps.

6. Response and task competition effects are additive with RSI. (Experiments 1, 3, and 4). In addition, response competition is (and task competition probably is) part of the mixed list cost not the shifting cost, i.e., it is no larger on AABB-different stimuli than on AABB-same stimuli (Experiment 3). That response and task competition are not affected by RSI makes sense according to the RSG model since the RSI is used to decide which task to do, and this decision is made before task specific processing begins (where competition effects presumably take place). One might have thought that since the subject becomes more "tuned" for a task by doing it, that response and task competition would be larger on AABB-different than AABB-same stimuli. That this does not occur is counter-intuitive but not inconsistent with the RSG model.

7. When a random sequence of univalent stimuli follow an alternating or non-alternating bivalent list a set expectation effect occurs (Experiment 7). A set expectation effect occurs because subjects must be set for one task or the other (they cannot be set for both tasks at once) and when they set themselves for the wrong task, they pay a time cost (in the form of re-setting themselves for the other task). Why is it not possible for subjects to be set for both tasks at once? Presumably, this is what occurs when the tasks are disjoint, and why there is no alternation cost in that case. It would seem that when some of the stimuli that subjects are presented with are bivalent, the subjects learn to have only one task in their current set at a time, and this cannot be overcome even when the stimuli are univalent.

8. Above (7), when the subject initially sets the response selection mechanism based on the preceding task sequence, there is no cost associated with setting it for a different task from the task before versus the same task (Experiment 7). The RSG model explains this by supposing that the decision of what task to get set for is performed

very quickly in this case (perhaps by some built-in sequence prediction mechanism that also acts in 2-choice serial RT) and takes no longer to make a decision to expect an alternation of task than a repetition of task.

And from the introduction:

9. When disjoint tasks are used there is no or little alternation cost. As described in connection with the full set switch hypothesis in the Introduction, with disjoint tasks subjects may be able to hold a set in which they can be ready for both tasks, and thus the response selection mechanism never needs to be re-set. In addition, task competition is not present with disjoint tasks, so this source of slowing (and with it perhaps criterion effects) is gone.

10. When task cues are appended to the stimuli the alternation cost is reduced. It was suggested in the introduction in connection with the full set switch hypothesis that task cues could reduce the alternation cost in one of two ways. One was by reducing the time needed by the subject to perform the set switch. The other was by allowing the subject to use the cue plus stimulus to select a response, thereby making the tasks disjoint. Either of these explanations could also apply here. The former is particularly congenial with the RSG model, since it is a decision that is made during the RSI, which could quite plausibly be sped up by a cue.

11. Executive PI. Same as full set switch hypothesis in the Introduction. In addition, it might be that task and response competition effects (from the now always irrelevant tasks) play a role in executive PI.

The RSG model naturally explains a large set of findings. However, it is useful to explore other possible models. Suppose one granted that there is both a baseline and an RSI component that each reflect a different problem involved in preparing for a task, and that the baseline component is basically as we have explained it, but that the RSI component is different than we have supposed. In the RSG model the RSI component reflects a set decision. We earlier considered that it might reflect some type of a switch process. However, a gradual set switch is inconsistent with the RSI benefit found when a univalent stimulus does not continue the task sequence (5). Furthermore, a discrete set switch in which the alternate task set is inhibited during the switch is ruled out by the set expectation effect found in Experiment 6 (7), since both task sets should be loaded on the univalent part of the list after inhibition has been turned off. Finally, if a set switch is used to load the task set for the upcoming task, it does not make sense that the switch can sometimes be performed without taking any time (8). Thus, the possibilities considered before are not viable alternatives. Perhaps there are other models that we have not considered that would also explain the data. It is not clear to us, however, what these models would be.

Perhaps it is the case that all we need is RSG1-RSG4, and that the idea of setting the mechanism for one task or another is not needed. According to this view, the RSI component of the alternation cost is the same thing as the non-alternating RSI effect, only it is increased by what might be described as "nuisance" factors (such as different criteria in the alternating and non-alternating conditions). The larger alternation cost for univalent stimuli that do not continue the bivalent sequence could also be explained in terms of these nuisance factors, and this would be consistent with RSI helping even when the univalent stimulus did not continue the bivalent sequence. However, surely something happens over the RSI that allows subjects to control which task they are about to perform, even if this process does not correspond to the whole

RSI benefit. But if there is some process that occurs during the RSI that takes only a little time and hence only accounts for a small part of the RSI benefit, why is there a large set expectation effect in Experiment 6 (7), that approximately corresponds to the size of the alternation cost, i.e. RT_{AAAB} - RT_{AAAA} is about the same size as the alternation cost on bivalent stimuli?

Finally, it might be that we are wrong in supposing that the speedup found at long RSI when a task is repeated on mixed lists (AABB-same vs. AABB-different stimuli) reflects a type of preparation that can only be changed by doing a task and not by "free time". However, we see no other way to explain why there is such a large advantage for AABB-same stimuli over AABB-different stimuli and why this advantage does not go away at the long RSI. For example, suppose that there was no "tuning" effect but a set decision does need to be made before performing a task. At 0 ms RSI the shifting cost occurs because the set decision needs to be made. However, at 400 ms RSI the shifting cost is still 150 ms, and this would have to correspond to the average amount of time over 400 ms that the set decision takes. Yet, increasing RSI from 200 to 1500 ms in Experiment 2 only resulted in another 78 ms speedup. The numbers just do not add up unless we suppose a "tuning" effect.

In summary, of all the possibilities we have considered the RSG model provides the best account of our rather complex set of data. Still, it is possible that there are other models that did not occur to us that can also explain the data. It would provide additional support for the RSG model if it generalized to different situations and helped explain other findings in the literature. Below we attempt to generalize our findings to a slightly different paradigm and in the Summary and Conclusion Section we will attempt to relate these findings to a much broader literature.

<u>The generality of the RSG model</u>

The RSG model does a good job of explaining this data, but it is intended as a general model of preparation and task set, so it is important to test it in situations other than alternating between two tasks. One such situation, albeit still in the laboratory, was investigated by Sudeven and Taylor (1987). In their paradigm subjects performed two tasks on digits: in the odd-even task subjects pressed one key if the digit was odd, and the other if it was even. In the hi-lo task subjects pressed one key if the digit was greater than 5 (hi) and the other if it was not (lo). Subjects knew which task to perform based on a cue that appeared a variable stimulus onset asynchrony (SOA) before the stimulus. There were 4 basic findings: 1) mean RT decreased as SOA increased up to SOAs of 2 seconds. 2) compatible stimuli (e.g., 3 was compatible because lo and odd were signalled on the same response key) were faster than incompatible stimuli. 3) After 5-10 days of practice the compatible stimuli were responded to roughly as quick with 0 ms SOA as with 2 second SOA. 4) Even after 17-20 days of practice there was still an SOA effect up to 2 seconds on the incompatible stimuli.

These findings are compatible with the RSG model. The effect of SOA up to 2 seconds is consistent with the finding that RSI effects last up to 200 ms, since the former includes the time to interpret the cue. Points 2 and 4 correspond to analogous findings with the alternating task paradigm (Jersild showed that practice did not eliminate the alternation cost). Although 3 has no straightforward interpretation in the RSG model, it is not inconsistent with it. In addition, the RSG model makes 2 untested predictions:

1) When the same task is performed throughout a block of trials (as in our non-alternating blocks), subjects should be much faster than when they get a large SOA between the cue and stimulus and the task order is random within the block of trials.

2) When the task order is random, there should be a beneficial effect of doing the same task twice in a row that persists even when a long SOA between the cue and stimulus is provided.

Note that these predictions do not necessarily follow from our present findings. If one thought, for example, that the results with the alternating task paradigm were caused by problems specific to keeping track of what task one is supposed to do next and not with task set in general, then there is no reason to predict that when subjects are given plenty of warning of which task is to be done next, that having done the same task last time will help at all, or that always doing the same task within a block of trials will be any help either. In Experiment 8 we test these predictions.

Experiment 8

In Experiment 8 we employ the odd/even and hi/low tasks of Sudevan and Taylor (1987). Stimuli were presented with a fixed RSI of 4000 ms. Task cues preceded the stimulus by an SOA of 0, 500, 1000, or 4000 ms. In addition to a random task order as used by Sudevan and Taylor, on some blocks of trials subjects performed the same task on each stimulus in the block and on others they alternated between the two tasks. In these two cases cues were still presented but were not necessary for doing the task.

Method

Subjects. Twenty-four undergraduate students at the University of California, San Diego participated for partial fulfillment of a course requirement.

Apparatus and Stimuli. The stimuli were the digits 2 through 9. The digits were colored white and were 1.2 cm high by 0.7 cm wide. There were two possible task cues: The hi/lo task cue consisted of the letters "H" and "L" 2 cm to the left and right, respectively, of the stimulus. The odd/even task cue consisted of the letters "0" and "E" 2 cm to the left and right of the stimulus. The letters in the task cues were 1.2 cm high and 0.7 cm wide, just like the digits.

The apparatus was the same as in Experiments 1-7. The "b" and "n" keys on the computer keyboard were used as the response keys. The "b" corresponded to hi and "n" to lo in the hi/lo task. The "b" and "n" corresponded to odd and even, respectively, in the odd/even task.

Tasks: Odd/Even Task: Subjects pressed the left response key if the stimulus digit was odd and the right key if it was even. Hi/lo task: Subjects pressed the left response key if the digit was greater than or equal to 6 and the right response key if it was less than 6.

Design. There were 5 block types employed in this experiment: 1) hi/lo task blocks, 2) odd/even task blocks, 3) alternating task blocks, hi/lo task first, 4) alternating task blocks, odd/even task first, 5) random order blocks. The order of the block types was as follows: The first 3 blocks were some random permutation of block types 1,3, and 5 or 2, 4, and 5. The next three blocks were some random permutation of the triple above not used in the first three blocks. This ordering of the first 6 blocks was repeated once more for a total of 12 blocks. Each block consisted of 40 trials. Stimulus digits were randomly selected with the constraint that no digit could occur on two consecutive trials. The interval between the cue and stimulus (stimulus onset asynchrony, or SOA) was randomly chosen from possible values of 0, 500, 1000, 2000, and 4000 ms for each trial.

Procedure. Subjects were given written instructions as before. The instructions emphasized that the subject should make all responses as fast and accurate as possible. After the subject had read the instructions the experimenter re-iterated them. Before the beginning of the experimental blocks subjects were given 6 practice blocks of 10 trials each. The block types of the practice blocks were determined just as the block types of the first 6 experimental blocks were determined. After practice the experimenter answered any questions the subject had about the procedure and then began the experiment.

Each block began with instructions displayed on the screen for 2000 ms describing the task sequence for that block. Before the first trial began, a white plus sign was displayed in the center of the screen as a fixation point for 1000 ms. The first trial began 500 ms after the fixation point was removed from the screen.

The stimulus for each trial was presented 4000 ms after the start of the trial, and the task cue was presented 0, 500, 1000, or 4000 ms before this, depending on the SOA for the trial. Subjects responded to the stimulus at this point, and RT was measured from the onset of the stimulus. If no response was made within 4000 ms or a response was made quicker than 200 ms a time-out error occurred, and this was treated just like an incorrect keypress. In the case of a correct response (that was not also a time-out error), the screen was cleared and the next trial began immediately. Thus, if the SOA for the next trial was 4000 ms, the stimulus for the current trial was replaced by the task cue for the next trial, otherwise, the screen was only cleared. In the case of an error the word "Error" was displayed in the center of the screen for 1500 ms, and the next trial began after a fixation point was presented just like before the first trial in the block.

Between blocks subjects were given the opportunity to rest while mean RT and the number of errors for each block were presented to them on the computer screen. When subjects were ready to continue they pressed a key and the experiment resumed.

Results and Discussion

Reaction Times. Figure 18A shows mean correct RT as a function of whether the sequence is fixed (alternating or non-alternating) or random, and whether the previous task was the same as the present one or not (always yes for non-alternating and always no for alternating) and SOA between the task cue and the stimulus. The first trial in each block and trials in which the stimulus repeated were not included in the analysis, nor were trials where the response was incorrect or trials that followed an incorrect response.

When the sequence was random there was a significant 250 ms SOA effect, and the overall effect of SOA was significant, $F(4,92) = 58$, $p < .01$. This replicates the finding of Sudevan and Taylor that more time with a task cue before the stimulus results in faster responses. There was also a significant 48 ms effect of repeating the last task, $F(1,23) = 20$, $p < .01$, and this did not interact with SOA, $F(4,92) = 1.2$, $p > .3$. The RSG model specifically predicts that even at long SOA, when the subject has 4 seconds warning of what task is to be done, there will be an effect of repeating the previous task. In fact, there was a significant 76 ms effect at 4000 ms SOA, $F(1,23) = 14$, $p < .01$.

The RSG model predicts that subjects will be faster on the non-alternating task blocks than the mixed blocks, even with a long SOA between the task cue and the stimulus on the mixed blocks. This prediction was met for random order blocks both when the task repeated, $F(1,23) = 11$, $p < .01$, and when it did not, $F(1,23) = 38$, $p < .01$. In addition, RT on non-alternating task blocks was 180 ms faster than RT on alternating task blocks, $F(1,23) = 100$, $p < .01$.

When the task sequence was fixed (alternating and non-alternating task blocks) the cue was not essential for performing the task. Nonetheless, there was a 171 ms effect of SOA, $F(4,92)=68$, $p <.01$. The SOA effect was 236 ms and 86 ms for alternating and non-alternating task blocks, respectively, and this difference was significant, F(4,92) = 9.2, p <.01. It might be that subjects find it difficult to keep track of the alternating sequence, and relied on the cue in this condition. This does not, however, explain the 86 ms effect when the sequence was non-alternating. The effect in this case, however, is likely to be a result of an artifact: at 0 ms SOA the cue and stimulus appear simultaneously, and subjects might not "notice" the appearance of the stimulus right away under these conditions.

Errors. Figure 18B shows error rate as a function of whether the sequence is fixed or random, and whether the previous task was the same as the present one or not and SOA. Of these three variables only the effect of the previous task was significant, F(1,23) = 9, p < .01, with 2.1% errors when the previous task was the same and 3.3% errors when it was different. Notice that the slower condition has the higher error rate, so a speed-accuracy tradeoff cannot explain the effect of the previous task on the speed of responding. The only other marginally significant effect was the interaction of fixed versus random sequence with whether the previous task was the same or different from the present one, F(1,23) = 4.2, .05 < p < .1. This appears to be because it is mostly on the fixed sequences where the effect of the previous task is found.

Conclusions

The results of Experiment 8 further support the RSG model. As predicted by the model, doing the same task as the time before leads to an advantage even with 4 seconds to get ready for the task. Moreover, when the same task was done on all trials in a block, subjects were much faster than in the other conditions.

The RSG model provides a viable explanation of our findings. If it is correct it has important implications for the machinery that is responsible for selecting responses and for how this machinery is programmed, as Allport and Styles (1992) put it, "to enable now one task to be performed, now another ". The model says there are two parts to being prepared for a task: 1) The response selection mechanism must be set for the right task, but before this a decision as to what task the mechanism is to perform must be made. Setting the mechanism itself takes little or no time but the decision might. 2) The mechanism varies in how ready ("tuned") it is for a task, and the only way to make the mechanism more ready for a task is to do the task (so the mechanism will now be ready if one does the task again).

Section Three: What are the limitations on the task sets that can be attained?

A common experience in activities such as athletics, driving, and video games is that when one is mentally "ready" for a particular event to occur, one is much quicker responding to it. A good example of this is playing defense in doubles volleyball. The goal of the defensive player is to be able to get to and "dig" any ball that the offensive player hits before it hits the ground. It is a common experience among advanced players that a "dink" (a short soft shot) is easy to dig if one is ready for it, but often very hard to react to and dig if one is not. The basic idea is that when one is ready for a particular shot one can react to it quicker and get to the spot on the sand where one needs to get to sooner. An obvious question is why players do not always make themselves ready for every shot they might see on a particular play.

There are at least two answers to this question. The first is very simple: experience. If the player does not know what shots to be prepared for he/she will obviously not be able to be prepared for them. Second, assume that the defender recognizes that in this particular situation there are three shots the offensive player can make. It is not necessarily the case that the defender will want to be ready for each of them. It might be that two of them are hard shots for the offensive player to make but unstoppable if made. In that case it would not make sense to prepare for those two shots if it would at all interfere with the defense of the third. The Hick-Hyman law, in fact, suggests that preparing for the two unstoppable shots would slow down the reaction to the third shot. The Hick-Hyman law (Hick, 1952; Hyman, 1953) states that in a choice-reaction time task (which is essentially what the defensive player is faced with) mean RT increases as the log of the number of alternatives (Hick, 1952; Hyman, 1953). Thus, in this situation it would not make sense for the player to be ready for all three shots, and instead the player should prepare for the one shot that he/she can stop.

In general, then, it makes sense for the player to: 1) evaluate what shots the offensive player might make in the given situation, 2) decide what subset of these shots to prepare for, and 3) become ready for the shots decided on in 2. Similar considerations apply to everyday activities in general. That is, people are in general quicker to respond to events they expect (are ready for). One might think that it would be best to remain ready for everything at all times. The reason this is not desirable is reflected in the Hick-Hyman law: the more one is ready for the slower one is responding to each possible event. Thus, becoming ready for some events can be seen as a way of trading off performance on the expected against performance on the unexpected.

This account, however, may be too simple. Point 3 assumes that players (people) can become ready for an arbitrary set of possible shots (S-R contingencies). This is challenged by anecdotal evidence. In particular, it is often observed that when defending against a player that one has never played against before and who can make shots that the defender has never faced, it is hard to be ready for both the new shots and the old shots for which the defender would normally be prepared. This is true even after the defender is aware of the new shots that the new player can make.

A similar phenomenon has been observed in the laboratory. Pashler and Baylis (1991a) practiced subjects on a choice-RT task with two stimuli mapped onto each of three keys. After 15 blocks of practice, two more stimuli were added to each key. On the 16th block RT to the old stimuli was slowed by over 200 ms, and was almost as slow as RT to the new stimuli. One reason that this might occur is the Hick-Hyman law. That is, subjects are prepared for a total of 6 S-R associations before block 16 and a total of 12 from then on. Thus, we would expect some slow-down of the old associations for this reason alone, although it is still hard to understand why subjects would be as slow with the old associations as with the new ones. In another experiment, however, they controlled for Hick-Hyman law effects. Instead of adding two stimuli to each response key, they replaced one of the stimuli with a new one. In this experiment they found again that RT to the old stimuli was slowed, this time by about 100 ms. Thus, beyond any Hick-Hyman law effect, adding a new S-R association to be prepared for disrupts performance of old S-R associations.

What do these laboratory results and anecdotal evidence from defense in volleyball suggest? People may be unable to choose any arbitrary set of S-R associations to prepare for. The Hick-Hyman law suggests that the total number of other S-R associations is important. The second of the Pashler and Baylis experiments discussed above suggests that whether the other S-R associations were learned around the time that the S-R association in question was learned also affects performance. In short, a subject's task set is not a simple sum of the currently relevant S-R associations.

In this section we look for limitations on the task sets which a subject can attain suggested by the above considerations and some of the results from the alternating task paradigm. The basic idea is that subjects might not be able to set themselves for just any combination of S-R associations, and the set switching processes discussed in connection with the RSG model, which we will review below, must be invoked when this happens. Thus, the results of Pashler and Baylis and the anecdotal evidence from defense in volleyball would be explained by the fact that subjects/players cannot be ready for the old and new stimuli/shots at the same time, and must spend some time switching between sets when they are incorrectly prepared.

We begin by reviewing the RSG model and some of the findings from the alternating task paradigm which seem relevant to the questions above. After this we will be ready to state the single mapping hypothesis, which states the ideas discussed in the preceding paragraph in a more precise way by drawing on the RSG model. After that, we will develop a plan for testing the single mapping hypothesis.

The RSG model

In Section Two we investigated performance when subjects alternated between two different tasks using bivalent stimuli. The results suggested the following conclusions about what must be done in order to prepare for a task. First, the response selection mechanism must be set for the task that is to be done. This can be likened to flipping a switch to one of (in this case) two possible positions, and does not itself take a large amount of time, although the decision of which way to flip the switch might take time. Second, the readiness of the mechanism for a task can vary, and readiness is increased only by performing a task. This can be likened to the re-programming of the internal machinery of the mechanism (and was termed the "tuning" effect). These two types of preparation were embodied in the RSG model.

In addition to the sources of slowing above, there were other sources of slowing when subjects alternated between two tasks. Task and response competition effects slowed subjects on alternating lists by around 120 ms, but only by about 15 ms on non-alternating lists. Thus, response and task competition account for upwards of 100 ms of the alternation cost. In addition, it is likely that some of the alternation cost is accounted for by a criterion effect.

This is when each of the stimuli in the list are relevant to both tasks (the stimuli are all bivalent). What about when the tasks are disjoint (the stimuli are all univalent) -- the primary focus in this section? In this case response and task competition effects (and perhaps criterion effects) should be gone. Thus, it makes sense that the alternation cost will be reduced. It is also possible that when the tasks are disjoint the process of setting the response selection mechanism for one task or the other will not take as long since the stimuli themselves may serve as task cues. However, it does not seem likely that the "tuning" cost is reduced just because the stimuli are univalent.

Thus, if the response selection machinery must be re-set when a different task from last time is performed, it should still be reflected in the alternation cost.

The fact is, Spector and Biederman (1976) found relatively little effect of alternation using disjoint tasks. What is more, in their Experiment 1 they found a 70 ms alternation benefit when there was preview (stimuli were in lists on sheets) and a 55 ms alternation cost when there was no preview (stimuli were on cards in a deck , one card at a time, self-paced). In addition, Jersild found a similar speed-up for mixed lists when the items were presented with preview (on sheets of paper)[11]. Even the no preview alternation cost of 55 ms is much smaller than the "tuning" effect found in Experiment 3 (upwards of 100 ms).

Why, then, is there such a small alternation cost with disjoint tasks? The most natural explanation (and the only one consistent with the RSG model) is that with disjoint tasks the subject is able to attain a set under which they are prepared for both tasks at once, and thus subjects do not have to re-set the response selection mechanism in order to prepare for each upcoming stimulus. The reason that this tactic does not work when the stimuli are bivalent, according to this account, is that if subjects were ready for both tasks on bivalent lists, response competition would slow the subjects down even more than switching, and errors would be hard to suppress.

So the findings of Spector and Biederman and Jersild with disjoint tasks suggests that subjects are able to become prepared for both tasks at once in this case, and as a result pay no cost of alternation. We found, on the other hand, that subjects could not become prepared for both the color and letter task in Experiment 6 of Section Two, even when the stimuli were univalent. In Experiment 6 subjects

---

[11] The interaction between alternation and preview may be caused by the fact that preview allows adjacent items on a list to interfere with each other, and the items interfere more when they are for the same task.

performed the color and letter tasks in alternating or non-alternating sequence on the first six items in a list (each of which was bivalent). The next four items in the list were univalent, and the task for the each stimulus (whether it was a color or a letter) was randomly determined for each item. It was found that even on the last univalent stimulus there was still a large cost associated with doing a different task from last time. Thus, even though there was no longer any threat of response or task competition, subjects were unable to be ready for both tasks at once.

The single mapping hypothesis

The reasons that our subjects in Experiment 6 were not able to become prepared for both tasks at once might indicate that they did not have a task set "available" to them that contained the S-R associations for both tasks. That subjects normally show no alternation cost (Jersild, 1927; Spector and Biederman, 1976) would seem to suggest that under normal circumstances they do have such a task set "available".

It might be, then, that the results from Pashler and Baylis and the contrasting pointed to above have a common explanation which we will call the single mapping hypothesis. This hypothesis holds that one result of practice is to "group" S-R associations into "mappings". In addition, at any given time only one mapping can be activated, and thus subjects can only be ready for two contingencies at the same time if they are part of the same mapping. Furthermore, the re-setting of the response selection mechanism described in the RSG model amounts to re-setting the mapping. Thus, the single mapping hypothesis claims that when subjects have to apply a different mapping than just before, a switching cost will be incurred. In turn, this means that when subjects perform a different task from just before they will incur a switching cost when the two tasks are held in separate mappings but not when they are held in the same mapping.

This hypothesis explains the Pashler and Baylis results by supposing that subjects cannot be ready for the old associations plus the new ones. Thus, subjects must learn a new mapping that contains the relevant old associations and the new associations, and they are initially slowed due to the lack of practice with the new mapping. A similar problem faces a defensive player in volleyball facing a new offensive player.

The single mapping hypothesis explains the fact that subjects generally show no alternation cost by supposing that these subjects learned to group the two tasks into the same mapping (hence, no need to ever switch the mapping on alternating lists). In Experiment 6, on the other hand, most of the stimuli are bivalent, and hence it would be detrimental to form a single mapping for both tasks. Thus, on the univalent part of the list subjects still only have one mapping for each task to switch into, and thus costs associated with switching are incurred.

The fact that Spector and Biederman's subjects seem to incur no cost of switching tasks whereas our subjects in Experiment 6 do incur one supports the single mapping hypothesis. However, there are several procedural differences which preclude any firm conclusions.

First of all, Spector and Biederman's subjects named the opposites of words and subtracted three from numbers, whereas our subjects performed an arbitrary 4-choice manual RT task in response to colors and letters. The difference may lie either in the fact that their tasks were both in some sense less arbitrary than ours, or in the fact that both of their responses were vocal (although Jersild found an alternation benefit with written responses).

Second, it might be that when the stimuli are bivalent subjects are punished (via errors) for trying to be ready for both tasks at once. For this reason, in Experiment 6 when subjects are responding on the univalent part of the list, they might be averse to being ready for both tasks. Thus, even if they could switch into a "combined" task set, they do not.

A third difference involves the method in which the cost of switching tasks was measured. In particular, Spector and Biederman and Jersild had subjects alternate between two tasks and compared the average time per item on alternating lists to that on pure lists. In our Experiment 6 the task sequence was random on the univalent part of the list and we assessed task switching costs with the cost of performing a different versus the same task as the time before. We consider the relationship between these two measures just below.

Task Repetition effect

According to the RSG model there should be a switching cost every time a different mapping is applied from the time before. Therefore, if two tasks are held in different mappings there should be a cost associated with performing a different task from the time before compared to performing the same task twice in a row. Spector and Biederman found little or no difference in the average RT on alternating lists and pure lists. As discussed above, since each stimulus on the alternating lists entails performing a different task from just before, the fact that there is no difference between alternating and pure list RT suggests that there is no switching required. A second way to measure costs of switching tasks is to give subjects lists in which the order of the tasks is completely random, and measure the task repetition effect (the effect on RT when the task but not the stimulus repeats). Superficially, these seem like comparable ways to investigate effects of switching task. However, they might not be comparable. Thus, it is important to know whether the lack of an alternation cost in Spector and Biederman (1976) is paralleled by a lack of a task repetition effect when the task order is random.

Spector and Biederman (1976, Experiment 2) had subjects name the opposites of words (subject says "good" in response to "bad") and subtract 3 from numbers (subjects say "12" in response to "15") in random order blocks and pure blocks. These are the same tasks used by them (and Jersild) to study the alternation cost with disjoint stimuli. They found that subjects were not significantly faster responding on pure blocks (a 35 ms non-significant advantage) and there was no task repetition effect (actual effect and F-value not reported). Thus, these findings parallel their finding of little or no alternation cost with the same tasks.

Other researchers, however, have found a task repetition effect (Forrin 1974; Marcel and Forrin, 1974; Rabbit and Yvas, 1973, also see Duncan, 1977). But the size of the effects are not large -- generally in the range of 15-60 ms. Thus, like the Spector and Biederman results, these findings show that the task repetition effect is very much smaller than the "tuning" costs with bivalent stimuli. Why there are task repetition effects at all in these cases will be considered in more detail in the General Discussion.

In short, the evidence cited above suggests that when two disjoint tasks are employed subjects show little or no cost of switching tasks. This is true whether the cost of switching tasks is measured with the alternation cost or with a task repetition effect using random order lists. According to the RSG model, then the response selection mechanism does not have to be re-set in these cases. According to the single mapping hypothesis, this further implies that both tasks are held in the same mapping in the above cases.

Basic Approach

Our basic approach for testing the single mapping hypothesis will be to test whether there are effects of the context in which two tasks were learned on whether there is a cost associated with following one task with the other one (a switching cost). The stimuli in the following experiments will all be univalent. Thus there will be no aversive effect of being in a set that includes both tasks as we suggested could be the case in Experiment 6. The logic here is as follows: If the learning context influences whether or not subjects show switching costs, then it is consistent with the single mapping hypothesis (including the assertion that switching costs reflect the need to re-set the response selection mechanism). In addition, these effects are naturally predicted by the hypothesis. Such effects are not, on the other hand, naturally predicted if the hypothesis is wrong.

We start by trying to test a simple working hypothesis, Working Hypothesis One: When two tasks are learned on mixed lists subjects will form a task mapping that includes both tasks. In contrast, when two tasks are learned on pure lists subjects will form one task mapping for each task.

Notice that this working hypothesis does not follow from the single mapping hypothesis. It merely embodies one guess about how subjects will group tasks into mappings under the assumption that the single mapping hypothesis is correct. This working hypothesis is tested in Experiment 10. Before that, however, we make sure that the basic finding that there is no alternation cost with disjoint tasks holds up when using arbitrary manual choice RT tasks as we do in this paper.

Experiment 9

In Experiment 9 we attempt to replicate Spector and Biederman's findings of no or little alternation with disjoint tasks, using arbitrary 4-choice manual RT tasks as were used in our Experiment 6. In addition, we attempt to replicate the effect of preview that they found on the alternation cost.

Methods

The methods used in Experiment 9 were identical to Experiment 2 except as noted.

Subjects. Twenty-four undergraduate students at the University of California, San Diego participated for partial fulfillment of a course requirement.

Apparatus and Stimuli. The stimuli in this experiment were colored discs (red, green, blue, and yellow) and white symbols (@,#,%,&). The discs were 1 cm X 1 cm and the symbols were 1.2 cm high X 0.7 cm wide. In the different keys condition the color task keys were oriented horizontally and symbol task keys vertically; in the same keys condition both tasks were performed on a horizontally arranged set of keys.

Tasks. Color Task: subjects pressed the response key corresponding to the color of the stimulus. Symbol Task: subjects pressed the response key corresponding to the symbol that is presented.

Stimulus Lists. Each list contained a total of 10 items. Whether each item was a color or a letter depended on the task called for by the sequence. The actual colors and symbols were randomly determined with the constraint that no response key would be the correct response on two items in a row in the same keys condition. This meant that red could not follow "@", for example, because those two stimuli corresponded to the same response key in the same keys condition, and this constraint held even in the different keys condition, where red and "@" corresponded to different (but corresponding) keys.

Design. There were sixteen blocks of 10 lists. Preview and task sequence were manipulated between blocks. There were four different task sequences, two alternating (color-symbol and symbol-color) and two non-alternating (color and symbol), for a total of eight total block types (2 preview X 4 sequences). The order of the block types was completely randomized. Twelve subjects were in the same keys condition and twelve subjects were in the different keys condition.

Results and Discussion

Mean correct RT on alternating lists was 759 ms, compared to 746 ms on non-alternating lists, and this difference was not significant, $F(1,23) = 1.3$, $p > .25$. This meager alternation cost replicates the main finding of Spector and Biederman (1976, Experiment 1). Spector and Biederman also found that providing preview produced a 70 ms alternation benefit. We found no interaction of alternation and preview, $F(1,23) = 1.6$, $p > .2$, although there was a trend in this direction: the alternation cost with no preview was 21 ms, whereas the alternation cost with preview was 4 ms. There was an overall 121 ms benefit of preview, and this was significant, $F(1,23) = 86$, $p < .01$. The mean by-item error rate was 2.0% (0.24 restarts per list) and there were no effects of alternation, preview, and no interaction of alternation and preview, $F<1$ in all cases. Thus, Spector and Biederman's main result that there is no or little alternation cost with disjoint tasks holds up when arbitrary manual choice RT tasks are used instead of their (and Jersild's) non-arbitrary naming and subtraction tasks.

Experiment 10

In Experiment 10 we investigate the effects of the context in which the subject initially encounters a task. The tasks were the color and letter tasks from previous experiments, except that the stimuli were always univalent. There were two phases of the experiment: the training phase consisted of 14 blocks of trials, followed by a test phase of 10 blocks. Half of the subjects were given only pure lists during the training phase, with no knowledge that they would eventually perform any other type of list or that other subjects performed any other type of list. Based on Working Hypothesis One, it was hypothesized that these subjects would form separate mappings for each task. The other half of the subjects performed lists with a random mixture of tasks (random order lists). Working hypothesis One holds that these subjects should form a single mapping that included both tasks, as appears to be the case when performing disjoint tasks on alternating lists. In the test phase of the experiment, both groups of subjects were given only random order lists. If the effect of learning the tasks in non-alternating lists is to form one mapping for each task, then the single mapping hypothesis predicts a task repetition effect during the test phase for these subjects. In addition, if the effect of learning the tasks in random order lists is to form a single mapping for both tasks, then the single mapping hypothesis predicts no task repetition effect for these subjects.

Method

Subjects. Sixteen undergraduate students at the University of California, San Diego participated for partial fulfillment of a course requirement.

Apparatus and Stimuli. The stimuli were the univalent stimuli from Experiments 5-7.

Tasks. The color and letter tasks from Experiments 3-7.

Design. The experiment consisted of two distinct phases : the training phase, consisting of 14 blocks, followed by the test phase, consisting of 10 blocks. Each block consisted of 50 trials. There were two groups of training subjects: the pure list training group and the random order training group. Within each training group four subjects were in the same keys condition and four were in the different keys condition.

Procedure. Subjects were given written followed by oral instructions as in previous experiments. Each training group was instructed on what they would be doing during the training blocks, with no mention that there would be any other blocks, how many blocks there would be, or that other subjects were doing anything differently. Before the training phase subjects practiced each task in a block of 20 trials (so these blocks were pure list blocks). This was done so that the experimenter could watch the subjects to make sure they understood the tasks before the experiment began. After the practice blocks the training phase began. When the training phase was over, a message was displayed on the screen asking the subject to get the experimenter. The experimenter then told the subjects what they would be doing during the test phase and started the subject on the test phase. Both phases of the experiment were conducted during the same one hour session.

The pure list training subjects performed the same task on each item in a block. The task for each block alternated between the color and letter task. Which task was done on the first block was randomly determined. The random order training subjects performed the color task if the stimulus was a colored disc and the letter task if the stimulus was a white letter. Whether the stimulus was a colored disc or a white letter was randomly determined for each stimulus. During the test phase, both groups of subjects were given only random order lists.

Each block began with instructions for what the subject would be doing. On non-alternating blocks this was "Color Task" or "Letter Task" depending on what task was to be done. On random order lists this was simply "Ready...". Before the first trial in each block a white plus sign was presented in the center of the screen as a fixation point. The fixation point was removed 1000 ms later, and 500 ms after that the first stimulus was presented.

Reaction time was measured from the onset of each stimulus. Responses faster than 200 ms or the lack of a response after 4000 ms were recorded as time-out trials. Immediately after a correct response was made to a stimulus, the next stimulus was presented. In the case of an error (including a time-out error) the word "Error" was presented in the center of the screen for 1500 ms and the next trial began after a fixation point was presented as on the first trial of the block. Thus, stimuli were presented with 0 ms RSI unless an error was made. One difference from Experiments 1-7 is that all stimuli are presented in the center of the screen, rather than offset to the right from the previous stimulus.

At the end of each block subject were given the opportunity to rest as mean RT and number of errors for each block were presented. When subjects were ready to continue they pressed a key.

Results and Discussion

Figure 19 shows mean correct RT during the training phase as a function of Block Number, and mean RT during the test phase as a function of Block Number and whether the previous task was the same or different (task repetition). The first trial in each block of trials was not included in the figure or the analyses that follow. In addition, when an error occurred, that trial as well as the next one were also not included.

Training Phase. Subjects were 115 ms faster at the end of training compared to the beginning of training, and the overall effect of block number during the training phase was significant, $F(6,72)=11$, $p < .01$. Pure list training subjects were a non-significant 29 ms faster than random order subjects, $F(1,12) = 1.9$, $p > .15$, and this factor did not interact with block number, $F < 1$.

As in some previous experiments there was a significant effect of key condition on RT, $F(1,12) = 11$, $p < .01$. However, unlike before subjects in the same keys condition were faster than subjects in the different keys condition (a 70 ms effect). In previous experiments it was assumed that the effect (in the opposite direction) was due to more response competition for the same keys condition than the different keys condition. In this experiment there is no response competition because the stimuli are all univalent. Thus, the present findings might indicate an advantage for the same keys condition that is also present in previous experiments but is cancelled out there by an opposite and larger effect due to response competition. Detailed consideration of the source of the key condition effect will be postponed until the General Discussion.

Test Phase. Random order training resulted in an average of 107 ms faster RT on the test phase than non-alternating training, $F(1,12) = 18$, $p < .01$, even though both groups of subjects were given random order blocks during this phase. This is consistent with other observations showing that more challenging practice can result in superior post-training performance (Schmidt and Bjork, 1992). In addition, this performance difference between training groups did not dissipate significantly over time, $F < 1$. Finally, Block Number did not have a significant effect on RT during the test phase, $F(1,48) < 1$, indicating that performance had reached an asymptote by this point.

The main purpose of this experiment was to test whether subjects who had learned the color and letter task in pure lists would have to perform a shift during the mixed lists that subjects who learned the tasks in this context would not have to do. If this is the case then there should be a task repetition effect for pure list training but not for random order training. There was an overall 102 ms benefit for repeating the task, $F(1,12)=84$, $p < .01$. There was no interaction, however, of task repetition and the type of training, $F<1$, indicating that repeating the task helped for both groups, and just as much for each. (There was also a significant 115 ms task repetition effect during the training phase for random order training subjects, $F(1,6) = 35$, $p <.01$). In addition, the task repetition effect found in the test phase did not significantly change over time, $F(4,48)=1.1$, $p > .35$, nor was the three-way interaction of task repetition, training, and block number significant, F<1. We will consider why both training groups had a task repetition effect in detail below.

As in the training phase, subjects were significantly faster in the same keys condition compared to the different keys condition, $F(1,12) = 26$, $p < .01$ (a 130 ms effect). In addition, task repetition and key condition interacted, $F(1,12) = 8.5$, $p < .05$, with a 69 ms task repetition effect in the same keys condition and a 134 ms effect in the different keys condition. There was also a marginally significant interaction of key condition and training group, $F(1,12) = 3.5$, $.05 < p < .1$, with key condition making a 177 ms difference for the pure list training group and only an 82 ms difference for the mixed list training group. The 3-way interaction between key condition, training group, and task repetition was not significant, F<1.

Errors. The error rates during the test phase for the pure list training group were 3.7% and 3.3% for when the task repeated and when it alternated, respectively. The corresponding error rates for the random order training group were 2.4% and 4.2%, respectively. Significantly more errors were made when the task was different from the trial before, $F(1,12) = 8.2$, $p < .05$. However, as can be seen from the numbers above, this was only true for the random order training group, and the interaction between training group and task repetition was significant, $F(1,12) = 21$, $p < .01$. The overall effect of training group on the error rates was not significant, F<1, and neither was key condition or any of its interactions with other variables.

Set expectancy curves

We have argued that when subjects need to switch task set when performing a different task from the one they performed on the stimulus before, there should be a task repetition effect. Indeed, the RSG model asserts that when subjects are in one task set they cannot put themselves into a new task set without actually doing the task once. However, it might be that there are other reasons that a task repetition effect occurs. For example, in the present experiment the stimulus sets for the two tasks are very distinct (colors versus letters). This might encourage subjects to anticipate whether the stimulus will be a color or a letter. When correct subjects might be quicker in the perceptual analysis of that stimulus. On the other hand, the effect might not be due to any "expectation" of which stimulus class (or task) will occur next, but will solely depend on a mismatch between the previous task and the current one.

These questions can be addressed using set expectancy curves. In order to make this clear we will consider some data from a 2-choice serial RT task. In a 2-choice serial RT task subjects are presented with one of two stimuli (say A and B) on each trial, and make one button press response if they see an A and a different one if they see a B. The next stimulus is presented some time after the subject makes a response to the previous stimulus. Figure 20A shows data from a 2-choice serial RT task with short RSI (50 ms from key up) and long RSI (50 ms from key down) as a function of the previous stimulus sequence from Vervaeck and Boer (1980, Experiments 1 and 2, respectively). These will be called stimulus expectancy curves. The sequences are laid out in the graph so that the recency and number of alternations in the sequence increases from left to right. For short RSI, RT increases from left to right. Thus, the more the recent sequence involves repeated elements, the faster RT to the next stimulus will be. Even if the sequence up to this point has been alternating, subjects are faster when the next stimulus is a repetition; and even if the task alternates, subjects are faster if just before the task repeated. However, with longer RSI's, during which subjects are presumably able to build up expectations for which stimulus will occur, RT first increases and then decreases from left to right. In short, the more the sequence resembles either alternating or non-alternating the more one is helped if the next stimulus continues this sequence and the more one is hurt if it does not.

Figure 20B shows the set expectancy curves in Experiment 6. These resemble the long RSI stimulus expectancy curves of Vervaeck and Boer (1980, Experiment 1). This suggests that subjects use the preceding task sequence to predict what task will occur next (even though the sequence has no predictive value) and commit to this task. When the preceding task sequence is AAA subjects appear to expect task A again most of the time. When the preceding task sequence is ABA subjects appear to expect task B most of the time. Nonetheless, subjects are worse off predicting and getting a task alternation (as for ABAB) than predicting and getting a task repetition (as for AAAA)[12]. This corresponds to the "tuning" cost. This difference, therefore, may be a more "pure" measure of switching costs than the task repetition effect, since the later may also reflect other factors (such as expectation of stimulus class).

Figure 21 shows the set expectancy curves for each of the training groups in Experiment 10. These set expectancy curves are qualitatively similar to what was found in Experiment 6. In particular, there is a 73 ms advantage of the AAAA sequence over the ABAB sequence. This suggests that just like in Experiment 6, subjects are unable to be prepared for both tasks at once. As a result, every time they have to perform a different task from the time before, they incur a "tuning" cost. This is true even when they "expect" the task to alternate, as is the case when the sequence is ABAB.

---

[12]It could be that subjects expect task A following AAA more often than task B following ABA, and this accounts for the difference in RT when the sequence is ABAB compared to AAAA. However, there is independent support for the "tuning" cost (Experiment 3), so it seems most natural to assume that the RT difference reflects this. If subjects expected task A following ABA on a large proportion of the trials then it follows that the difference in RT when the sequence is ABAB compared to AAAA should be much larger than the "tuning" cost, but it is not. Finally, Experiment 11 will provide evidence that subjects expect task A after ABA no more often than they expect task B after AAA.

Why is there a task repetition effect for both training groups?

Spector and Biederman (1976, Experiment 2) found no task repetition effect when stimuli were presented in random order to subjects from the start of the experiment. This is similar to our random order training group. However, we found a large task repetition effect in this case. There were several procedural differences which may have played a role in the difference. First, we use two 4-choice RT tasks in which the stimuli are arbitrarily mapped onto one of four response keys. The tasks used by Spector and Biederman, on the other hand, were to subtract three from numbers and name the common opposites of words. Both of these tasks are in some sense less arbitrary than the tasks we used. In addition, both of their tasks require vocal responses and both of ours require manual responses. Yet another difference between their procedures and ours is that they had an approximately 4 second RSI whereas we use a 0 ms RSI. Any of the above factors may account for the difference between the findings of Spector and Biederman and our own. These factors do not seem to matter, however, for the alternation cost, since we found no alternation cost in Experiment 9 with procedures similar to the ones employed in Experiment 10.

Why, then, is there a task repetition effect for both of our training groups? What does this mean for the single mapping hypothesis? If both the single mapping hypothesis and Working Hypothesis One were correct, then there should have been a task repetition effect. Thus, one of the two is wrong. It might be, then, that the single mapping hypothesis is wrong, and that with the color and letter tasks that we used and the particular procedures we used, subjects will always show a task repetition effect no matter what the context in which they learn the tasks, but will not show an alternation cost. With the tasks used by Spector and Biederman, however, subjects will show neither an alternation cost nor a task repetition effect when the task order is random.

On the other hand, the single mapping hypothesis might be correct and Working Hypothesis One is wrong. That is, it is not the case that mixed list training will result in subjects forming a single mapping containing both tasks was wrong. If the single mapping hypothesis is to explain why there is no alternation cost in Experiment 9, then it would have to be the case that alternating list training results in a single mapping for both tasks. Thus, we can revise our working hypothesis as follows, Working Hypothesis Two: With pure list training and random order training subjects will form separate mappings for each task; with alternating list training subjects will form a single mapping that contains both tasks.

The single mapping hypothesis and Working Hypothesis Two therefore make two predictions: 1) when subjects are given alternating list training rather than pure list or random order list training, they should show no task repetition effect during the (random order) test phase. 2) when subjects are given random order or pure list training, they should show an alternation cost when the test phase is alternating. These two predictions are tested in Experiments 10 and 11, respectively.

Experiment 11

In Experiment 11 subjects perform the color and letter tasks on alternating lists during the training phase. Then, during the test phase, subjects are given random order lists as in Experiment 10. This experiment provides a strong test of the single mapping hypothesis. If this hypotheses is true, then it must be the case that random order and pure list training (with our tasks) results in the subjects forming one mapping for each task. This is because there were task repetition effects during the test phase for both training groups in Experiment 10. It must also be that alternating list training results in the formation of a single mapping that contains both tasks. This is because there is no alternation cost in Experiment 9. Thus, a task repetition effect in the test phase of this experiment can prove the single mapping hypothesis wrong. In short, this experiment tests whether the single mapping hypothesis and Working Hypothesis Two are both correct. But due to the results of Experiment 9 and 10 the single mapping hypothesis cannot be true unless Working Hypothesis Two is also correct. Therefore this experiment really tests the single mapping hypothesis.

Method

The methods used in Experiment 11 were identical to Experiment 10 except as noted.

Subjects. Twelve undergraduate students at the University of California, San Diego participated for partial fulfillment of a course requirement.

Design. There was only one training group: the alternating task training group. Six subjects were in the same keys condition and six were in the different keys condition.

Procedure. The only procedural difference from Experiment 10 was that subjects were only given alternating task lists during the training phase. The instructions at the beginning of each block of the training phase were either "Color -- Letter" or "Letter -- Color", depending on which task was done first on that block. Subjects alternated between color-letter and letter-color blocks; it was randomly determined which order was done on the first block.

Results and Discussion

Figure 22 shows mean correct RT during the training phase as a function of Block Number, and RT during the test phase as a function of Block Number and task repetition. Trials on which an error occurred, trials following an error, and the first trial in each block were not included in the figure or the analyses that follow.

Subjects were an average of 103 ms faster at the end of training compared to at the beginning of training, and the overall effect of block during the training phase was significant, $F(6,60) = 10.2$, $p < .01$. Performance appears to have reached an asymptote after this since the effect of block during the test phase was not significant, $F(4,40) = 1.4$, $p > .25$.

Test Phase. In contrast to the previous experiments there was no significant task repetition effect, $F(1,10) = 3.9$, $.05 < p < .1$, and this did not change with block number, $F<1$. There was, it should be noted, a consistent trend across blocks in favor of doing the same task as last time of 20 ms. However, this trend is much smaller than the 102 ms effect found in Experiment 10. Thus, although it may be too strong to claim that there is no task repetition effect at all, whatever is going on in this experiment seems to be much different than what occurred in Experiment 10. In particular, these findings support the hypothesis that subjects do not have to re-set the response selection mechanism when performing a different task from last time, as appears to be the case with pure list and random order training.

The overall mean RT for subjects in the same keys condition was 71 ms faster than for subjects in the different keys condition. This trend is in the same direction as the significant trend from the previous experiment. However, the difference was not significant here, $F<1$, nor were any interactions of key condition with other variables.

Error Rates. The error rate during the test phase was 3.4% when the task repeated and 2.7% when it did not. This difference was not significant, $F(1,10) = 3.1$, $p > .1$. Key condition had no significant effects or interactions on the error rates.

Set expectancy curves

The set expectancy curve for Experiment 11, shown in Figure 23, further supports the hypothesis that there is no need to perform some type of switch when a different task from last time is performed. There are two major differences between this curve and those from Experiments 9. First, this curve is shallower: subjects were only 99 ms faster when the sequence was AAAA than when it was AAAB, compared to a 147 ms difference in Experiment 10. Second, subjects were no slower responding to the ABAB sequence than the AAAA sequence. In Experiment 10 subjects were 72 ms faster when the sequence was AAAA compared to ABAB. This is very important. We have argued that when the subjects are presented with a task sequence of ABA they expect task B to occur next and hence ABAB corresponds to the case where they expect a different task and get it. Similarly, AAAA corresponds to a situation in which subjects expect the same task again and get it. That the later condition is no faster than the former suggests that there is no inherent cost in doing a different task from last time -- the response selection mechanism is ready for both tasks at once.

Nonetheless, there does appear to be some type of expectation effect since, otherwise, the curve in Figure 23 should have been flat. This expectation effect, then, explains the small task repetition effect that is present. That is, on the whole subjects expect repetition of task somewhat more often than alternation.

But if subjects have a single mapping that includes both tasks, why is there an expectation effect at all? It might be, contrary to our arguments above, that subjects can still only be ready for one task at a time. That subjects are just as fast when the sequence is ABAB as when it is AAAA might have other explanations that we have not considered. However, it is also possible that our arguments above are correct, and subjects are ready for both tasks at once, but that they commit to one task or the other for other reasons. For one, it might be that there is a cost of committing to the wrong task even if there is no benefit to committing to the correct task, and subjects commit to one task or the other based on the sequence simply because of the belief that they can predict what the next task will be, i.e., "gamblers fallacy". It might also be that subjects can prime some S-R associations at the expense of others within the same mapping, a "within mapping" expectancy effect. Finally, in the different keys condition subjects may be able to prime response hand. Indeed, in Experiment 10 the task repetition effect was larger in the different keys case, consistent with this hypothesis, and the set expectation curves are "deeper" in the different keys case in both Experiment 10 and 10. Thus, there are reasons why there might be set expectation effects other than the need to switch set that is suggested by the RSG model. However, the set switching process of the RSG model necessarily predict a difference in the speed of responding when the sequence is AAAA compared to ABAB, and this difference is not present here. Thus, these data suggest that the response selection mechanism does not necessarily need to be re-set after doing a different task from the task before when the two tasks are learned on alternating lists.

Summary

Experiment 11 supports the single mapping hypothesis (with Working Hypothesis Two). In particular, the prediction that alternating list training would result in no task repetition effect during the test phase was met. In addition, the set expectancy curves further suggest that there is no "tuning" cost at all. Thus, with alternating list training, subjects are able to be ready for both tasks at once. According to the single mapping hypothesis subjects learn a mapping that includes both tasks. In Experiment 12 we test the second of the predictions of the single mapping hypothesis.

Experiment 12

In Experiment 12 subjects were given both pure and mixed list blocks. Three different training groups were given three different types of mixed lists during the training phase. One group was given random order lists, one group was given alternating lists, and the third group was given AABB lists. During the test phase, each of the groups performed mixed and pure lists as before, but now the mixed lists were alternating lists for all of the groups. In Experiment 9 we found little or no alternation cost when subjects performed alternating and non-alternating (pure lists) task sequences throughout the experiment. This corresponds to the alternating list training group here, so we expect little or no alternation cost for this group during the test phase.

According to the single mapping hypothesis, there was a task repetition effect for both pure and random order list training groups in Experiment 10 because subjects in these groups formed separate task mappings for each task. The single mapping hypothesis, then, predicts that our random order list training group in Experiment 12 should do the same, and as a result there should be a large alternation cost during the test phase.

Why would it be the case, as suggested by Experiments 9 and 10, that learning two tasks on alternating lists leads subjects to form one task mapping including both tasks, but learning two tasks on random order lists leads subjects to form a separate mapping for each task? Alternating lists differ from random order lists in two ways. First, the sequence is fixed so that the subject always knows what the next task is. Second, the subject performs the same task as last time every time on alternating lists but only half the time on random order lists. The AABB lists resemble alternating lists on the first point and random order lists on the second. Thus, whether or not AABB list training results in a task repetition effect will identify which of the two factors is the determining factor.

Methods

The methods used in Experiment 12 were identical to Experiment 10 except as noted.

Subjects. Thirty-six undergraduate students at the University of California, San Diego participated for partial fulfillment of a course requirement.

Design. There were three training groups: random order, AABB list, and alternating list training groups. For each training group six subjects were in the same keys condition and six were in the different keys condition.

The experiment consisted of a total of 28 blocks of 50 trials each. The training phase lasted the first 16 blocks, the test phase the remaining 12. Each subject was given both pure list blocks and mixed list blocks. During training the mixed list blocks were random order blocks for random order subjects, AABB list blocks for AABB list subjects, and alternating list blocks for alternating list subjects. During the test phase the mixed blocks were alternating list blocks for all subjects. The block types of the first four blocks were randomly determined with the following two constraints. First, block type alternated between mixed and pure lists. Second, each of the two versions of the pure lists (pure color or pure letter) occurred once in these four blocks, and each version of the mixed lists occurred once, except for random order list subjects, since there is only one version of random order lists: random. This ordering of the block types was then repeated six more times for a of 28 blocks, except that during the test phase the mixed lists were now alternating lists (as described above).

Results and Discussion

Training Phase. Figure 24 shows mean correct RT during the training phase for each of the three training groups as a function of pure versus mixed list block and Block Number. The first trial in each block, trials in which an error occurred, and trials following an error trial were not included in the figure or the analyses that follow.

Subjects were 116 ms faster at the end of training than at the beginning of training, $F(3,90) = 53$, $p < .01$. This practice effect did not differ significantly between groups, $F<1$. In addition, the type of training did not significantly effect overall RT, $F<1$, even though the composition of the mixed lists varied depending on the type of training.

In Experiment 9 we found only a 21 ms alternation cost with disjoint tasks when, as here, no preview of the next stimulus was allowed. The alternation cost found during the training phase for the alternating list training group was 93 ms. Though larger than what we found in Experiment 9, it was smaller than the effect of pure vs mixed list for the other two training groups: 118 ms and 171 ms for random order AABB list training groups, respectively. The effect of training on the pure vs mixed list cost was significant, $F(2,30) = 4.6$, $p < .05$, as was the overall effect of pure vs mixed list, $F(1,30) = 144$, $p < .01$.

The error rates for random order training subjects were 2.4% and 2.6% on pure and mixed lists, respectively; the corresponding error rates for AABB list training subjects were 3.8% and 4.3%, respectively. Thus, for both of these groups there were more errors in the mixed list condition, particularly so for the AABB list training group. A speed-accuracy tradeoff, therefore, cannot explain the slowing on mixed lists. The alternating list training group, on the other hand, had a higher error rate on pure lists than on mixed lists (3.1% versus 2.6%), and the corresponding interaction between training group and list type was significant, $F(2,30) = 4.0$, $p < .05$. Therefore, a speed-accuracy tradeoff may explain the larger than expected alternation cost for the alternating list training group. Overall, there were not significantly more or less errors on pure lists than on mixed lists, $F<1$.

It should be emphasized that for both AABB and random order lists, the subject does the same task as last time on half the stimuli, whereas for alternating lists the subject does the same task as last time on every stimulus. Thus, that the mixed versus pure list effect is smallest for the alternating list training group clearly shows that these subjects have a smaller cost associated with doing a different task from just before than the other two training groups.

Key condition had no overall effect, $F(1,30) = 1.7$, $p > .2$, but did have several interactions. Key condition interacted with training group, $F(2,30) = 3.7$, $p < .05$, corresponding to the fact that the AABB lists were responded to slowest overall in the same keys condition but fastest overall in the different keys condition. Overall, practice effects were smaller in the different keys condition, as indicated by the interaction between key condition and block number, $F(1,30) = 3.2$, $p < .05$. This was only true for the alternating and AABB sequences, however, with Block Number having a slightly larger effect in the different keys condition (as was the case in Experiment 10). The 3-way interaction between key condition, training group, and Block Number was significant, $F(1,30) = 2.3$, $p < .05$.

Task repetition effects during training. Above we argued that AABB lists were like random order lists in that the task repeats on these lists half the time, but like alternating lists in that the task sequence is fixed. Thus, supposing for the moment that the single mapping hypothesis is correct, whether there is a task repetition effect on AABB lists identifies which of these two factors determines whether subjects will form one mapping for both tasks (as appears to be the case on alternating lists) or form separate mappings for each task (as appears to be the case for random order lists).

Figure 25 shows mean correct RT during the training phase for the random order and AABB list training groups as a function of Block Number and whether the previous task was the same or different (the alternating list training group is excluded from this analysis since the previous task is always different from the present task on alternating lists). There was an overall task repetition of 137 ms, and this was significant, $F(1,20) = 50$, $p < .01$, and it did not interact with the type of training, $F(1,20) = 1.0$ (115 ms and 159 ms effect for random order training and AABB list training, respectively). The error rate when the task alternated was higher than when it repeated, $F(1,20) = 5.7$, $p < .05$ (3.8% versus 3.1%), so it cannot be that the task repetition effect is a speed-accuracy tradeoff. The task repetition effect was 155 ms in blocks 1-4, 106 ms in blocks 5-9, 170 ms in blocks 10-14, and 118 ms in blocks 15-18. As is apparent from these numbers, the task repetition effect did not diminish appreciably over the blocks. Nonetheless, the effect of block number on the task repetition effect was significant, $F(3,60) = 4.1$, $p < .05$. Finally, the 3-way interaction of training group, task repetition, and block number was not significant, $F<1$.

Since there is a task repetition effect on AABB lists it appears that the crucial factor is that the task repeats half the time and alternates half the time, and not the predictability of the sequence. (Unfortunately, we cannot test the other intermediate between random order lists and alternating lists: a task sequence that is unpredictable yet alternates every time).

In summary, two major conclusions follow from the results of the training phase. First, alternating list training results in a smaller cost of switching task than random order and AABB list training. Although alternating list subjects showed a larger than expected alternation cost, it was smaller than the effect of mixed vs pure list found on the other two training groups. Furthermore, a speed-accuracy tradeoff might explain why the alternation cost was as large as it was. Second, the reason that random order training results in a task repetition effect (Experiment 10) but alternating list training results in none (Experiment 11), is that on random order lists the task sometimes repeats and sometimes alternates, but on alternating lists the task always alternates. If the determining factor had been the unpredictability of random order lists then AABB list subjects should not have shown a task repetition effect, but they did.

Test Phase. During the test phase of Experiment 12, each training group was given pure and mixed lists, as before, but now the mixed lists were alternating for all three groups. The single mapping hypothesis predicts only a small alternation cost for alternating training but a large effect for random order training and AABB list training (because there was a task repetition effect during the training phase). Figure 26 shows mean correct RT in the test phase for each of the training groups as a function of alternation and Block Number. The overall alternation cost was 73 ms, and this was significant, $F(1,30) = 75$, $p < .01$. The size of this effect, however, depended on the training group. For alternating training subjects the effect was 37 ms (14-60 ms forms a 95% confidence interval on the effect). This is smaller than the size of the effect found by Spector and Biederman in the no preview condition. The alternation cost, however, was 82 ms and 102 ms, for the random order and AABB list training groups (46-118 ms and 68-136 ms form 95% confidence intervals around the respective effects). The interaction of training group and alternation was significant, $F(2,30) = 5.1$, $p < .05$. Training did not have a significant overall effect on RT during the test phase, $F<1$. Thus, these data are consistent with the single mapping hypothesis.

Unlike in Experiment 10 and 11, Block Number had a marginally significant effect during the test phase, $F(2,60) = 3.1$, $.05 < p < .1$. This effect was significantly larger on the alternating lists than the pure lists, $F(2,60) = 4.5$, $p < .05$ (30 ms difference between RT on block number 5 and 7 for alternating lists but only a 4 ms different for pure lists). In other words, the alternation cost decreased somewhat across blocks. Furthermore, the interaction of block number and training group was significant, $F(4,60) = 2.8$, $p < .05$, corresponding to the fact that the block number effect was virtually absent for the alternating list subjects. The effects of Block Number are consistent with the idea that some subjects in the random order and AABB list conditions learn, over the course of the test phase, how to perform on alternating lists without incurring a switching cost. According to the single mapping hypothesis, this means that some subjects learn to "merge" the mappings for the two tasks. This explanation would predict a 3-way interaction between block number, training group, and alternation, which was not significant, $F<1$. However, this would seem to be due to a lack of power since the predicted interaction is present.

As in the training phase, key condition had no overall effect, $F<1$, but did interact with several variables. As before, key condition and training group interacted, $F(2,30) = 4.4$, $p < .05$, corresponding to the fact that AABB subjects were the slowest of the three training groups in the same keys condition but the fastest of the three groups on the different keys condition. In addition, the 3-way interaction of key condition, training group, and alternation was significant, $F(2,30) = 4.3$, $p < .05$, corresponding to the fact that the alternating cost was larger in the same keys condition for AABB list training subjects but smaller in this condition for random order subjects. None of the other five interactions of key condition with other variables were significant.

The error rates on pure and alternating lists were 2.6% and 2.4%, 3.4% and 3.0%, and 2.2% and 2.1% for random order, AABB list, and alternating list training groups, respectively. There were no significant effects or interactions on the error rates (largest p-value > .2).

Summary

Experiment 12 further supports the single mapping hypothesis (Working Hypothesis 2). In particular, the prediction that there would be a large alternation cost with random order training but not with alternating list training was met. In addition, since AABB list training also led to a task repetition effect and a large alternation cost, it appears that the critical factor in determining whether subjects will learn to group two tasks into a single mapping or keep them in separate mappings is whether the task always alternates (as on alternating lists) or sometimes does not (as on random order and AABB lists and pure lists).

General Discussion

In Section Three we have tried to answer the question of why subjects sometimes appear able to prepare for two sets of S-R associations at the same time, and sometimes appear able to only prepare for one of the two at once. To explain this we proposed the single mapping hypothesis. This hypothesis assumes that when learning to perform a task subjects "group" different S-R associations into mappings. Under some circumstances two S-R associations will be grouped into the same mapping (for example, if they are from the same task) and other times they will be grouped into different mappings (for example, if they are from two different tasks involving the same set of bivalent stimuli). The critical assumption is that the response selection mechanism can only be "set" for one mapping at a time, where "set" is applied as in the RSG model. From this it follows that whenever two associations are learned in separate mappings there should be a cost associated with applying one association followed by the other, corresponding to the "tuning" of the response selection mechanism for the later mapping.

This hypothesis explains why there is no alternation cost with disjoint stimuli, yet there was a task repetition effect on the univalent part of the lists in Experiment 6. That is, when subjects learn two disjoint tasks on alternating (or alternating plus pure) lists, they form a single mapping that holds both tasks. Thus, on alternating lists, there is no alternation cost. When the stimuli are sometimes bivalent, as in Experiment 6, subjects learn to put the two tasks into separate mappings (in order to avoid errors and response competition). Thus, on the univalent part of the list, since they can only be set for one mapping at a time, subjects must switch mappings every time the task changes, and as a result a task repetition effect occurs.

The single mapping hypothesis also explains why adding one more shot to "worry" about in doubles volleyball, or one more S-R association to a choice reaction time task (Pashler and Baylis, 1991, Experiment 3-4), has adverse effects even on already well learned associations. That is, since subjects cannot be ready for both the old mapping plus the new associations, they must learn a new mapping (or alternatively form a new mapping containing the new associations and switch between the new and old mapping, yet again incurring a cost).

We tested the single mapping hypothesis by manipulating the context in which the subject learns two different tasks: on pure lists, on alternating lists, on random order lists, or on AABB lists. According to the single mapping hypothesis the learning context might affect whether the tasks are placed in the same mapping or in two different mappings. This, in turn, should affect whether or not an alternation cost occurs on alternating lists and whether a task repetition effect occurs on random order lists. Thus, the single mapping hypothesis predicts that the learning context should affect whether or not task repetition effects and alternation costs occur with two particular tasks.

In Experiment 9 we replicated the findings of Spector and Biederman (1976, Experiment 2) of no or little alternation cost with disjoint tasks, using arbitrary 4-choice manual RT tasks. Thus, as suggested above, the single mapping hypothesis would assert that subjects form a single mapping that holds both tasks. In Experiment 10 we found that both pure list and random order list training resulted in a task repetition effect during the random order test phase. The single mapping hypothesis, then, would hold that subjects form separate mappings for both tasks for both training groups.

Thus, the single mapping hypothesis makes two predictions. 1) With alternating list training, subjects should show no task repetition effect on random order lists. This prediction was met in Experiment 11. 2) Random order training should lead to a sizeable alternation cost (RT on alternating lists minus RT on pure lists) but alternating list training should not. In Experiment 12 the type of training did effect the size of the alternation cost. Random order training resulted in an 82 ms alternation cost whereas alternating list training resulted in only a 37 ms alternation cost.

<u>When do subjects form one mapping and when do they form two?</u>

Supposing, then, that the single mapping hypothesis is correct, what are the conditions that lead to subjects forming separate mappings for two tasks as opposed to one for both. Both random order training and pure list training appear to lead to the formation of two mappings, whereas alternating list training appears to lead to the formation of just one. This contrasts with our initial working hypothesis that mixed list training would lead to one and pure list training to two. Perhaps, then, mixed list training will result in one mapping when the sequence is predictable (as on alternating lists) but two when the sequence is unpredictable (as on random order lists).

This possibility, however, is countered by the findings with AABB list training in Experiment 12. Subjects with AABB list training showed a 159 ms task repetition effect during training, and a 102 ms alternation cost on the alternating test lists. Thus, AABB list training leads to the formation of one mapping for each task, even though the sequence is predictable. It appears, then, that with our tasks and procedures subjects form two mappings with pure list training and with mixed list training when the task does not always alternate (as is the case on random order lists and AABB lists), and a single mapping on mixed lists when the task always alternates (alternating list training). Why is this the case?

The answer may have to do with the Hick-Hyman law. That is, when subjects have two mappings they incur a switching cost every time the task changes. In addition, there is a cost associated with the number of S-R associations in a given mapping that is incurred on every stimulus (different task from last time or not) that corresponds to the Hick-Hyman law. Thus, if the former cost is larger, it makes sense to keep the tasks in a single mapping when the task alternates every time, but in separate mappings when the task sometimes alternates and sometimes does not. In particular, according to this account, it would be more efficient to keep the two tasks in separate mappings whenever the ratio of the Hick-Hyman cost to the switching cost is greater than the proportion of trials on which the task alternates.

Arbitrary tasks vs. non-arbitrary tasks

Using two arbitrarily 4-choice manual RT tasks we found that with random order training subjects show a task repetition effect on random order lists. Spector and Biederman (1976, Experiment 2), however, found no task repetition effect when the tasks were subtracting three from numbers and naming the opposites of words. As noted above, there are several differences between their experiments and our own. However, the most notable would seem to be that their tasks are less arbitrarily than ours. I.e., they could instruct their subjects on what to do with a simple rule ("name the common opposites of the words we show you") whereas we necessarily have to list each S-R association.

Other researchers who have found only small task repetition effects also used non-arbitrary tasks. Marcel and Forrin (1974) had subjects name letters and digits and found roughly a 45 ms task repetition effect. Rabbit and Yves (1973) had subjects in one group make spatially compatible button press responses to letters and digits. Thus, the numbers 1 through 4 were mapped onto fingers from left to right, and the letters A through D were mapped onto other fingers from left to right (the fingers for the two tasks were actually inter-leaved). These subjects showed a 46 ms task repetition effect. A second group of subjects made spatially compatible responses to numbers, as before, and also made spatially compatible responses to which of four neon lamps would illuminate. These subjects showed a 64 ms task repetition effect. Thus, although significant task repetition effects are found in these studies with non-arbitrary tasks, the effects are smaller than the task repetition effects we find in Experiment 10 and 11 using arbitrary tasks with random order training (115 ms in both cases).

Forrin (1974) employed one arbitrary task and one non-arbitrary task. He had subjects name digits and make a verbal digit response to shapes (an arbitrary task). He found task repetition effects of under 30 ms, although his data was pooled from four sessions. It might be that as long as there is only one or no arbitrarily tasks the task repetition effect will be small. Another reason that Forrin's (1974) subjects might not have shown a larger task repetition effect is that the data are averaged over four sessions. Thus, it might be that over the last two or three sessions the task repetition effect was reduced due to practice, and the overall task repetition effect is thus small.

Assuming for the moment, then, that the reason we found a task repetition effect and Spector and Biederman and these other researchers found none (Spector and Biederman) or only small ones is that our tasks were arbitrary and theirs were not, what implications does this have? One possibility is that only one arbitrary mapping can be set at a time, but non-arbitrary tasks can be independently added to the current task set without problems. An important question, if this is correct, is what makes a task arbitrary. Two possibilities seem most natural. The first is that after a certain amount of practice a task becomes non-arbitrary. The second possibility is that the task must be described by a simple rule.

Although interesting, these speculations on the role of arbitrariness are a bit premature. Perhaps it should be emphasized that although the task repetition effects found when non-arbitrary tasks are involved are much smaller than the effect found with arbitrary tasks, these are still reliable effects. It might be that the effects that are found are due to reasons separate from switching mappings. However, it might also be that the task repetition effects found in these experiments do reflect the "tuning" cost associated with switching mappings, but this "tuning" cost is smaller in these cases. It seems that set expectancy curves might help to decide this issue. In short, at this point it appears that the single mapping rule might only apply to arbitrary tasks, but it is still too early to tell for sure.

Other accounts

The single mapping hypothesis does a good job explaining this data. But perhaps there are other explanations. For example, in Experiment 12 the alternating list training subjects showed the smallest alternation cost. Perhaps this was due to the fact that since they learned the tasks on alternating lists they were somehow more efficient with that sequence, perhaps in the speed at which they retrieved what the next task was going to be and this somehow speeds them up. Thus, alternating list training subjects still show a cost with this sequence but the cost is smaller due to practice. This would explain why the alternation cost is larger during the training phase than during the test phase for these subjects. This cannot, however, explain why alternating list training subjects show a smaller task repetition effect on random order lists than random order training subjects do.

Still, there may be other possible ways to explain this data. Nonetheless, the virtue of the single mapping hypothesis lies in its ability to explain the pattern of alternation costs and task repetition effects in a very simple and intuitive way.

Conclusions

We began with some ideas about defense in doubles volleyball. In particular, the idea that having to be prepared for one additional shot causes problems in a players ability to also remain ready for shots the defender is used to defending against. The single mapping hypothesis would hold that the difficulties stem from the fact that the defender can become set for the old shots or the new shots, but not both at once. If the single mapping hypothesis is correct, it would also have important implications outside of defense in volleyball. Indeed, it might have practical implications for teaching skills, such as whether two skills should be taught alone or together (perhaps in alternation).

Summary

When subjects alternate between two tasks on lists of stimuli in which each stimulus is relevant to both tasks (bivalent stimuli) they show a large cost in the rate of responding compared to when they perform the same task on each item in the list. The experiments in Section One suggest that there is one component (the RSI component) of this cost that is overcome if the subject is given time to get ready for each upcoming stimulus. A second component (the baseline component) is not reduced, or at least reduced to a much smaller extent, when time is provided to prepare for the each upcoming stimulus.

The experiments in Section Two led to a more refined understanding of what is going on. First of all, Experiment 3 showed that there is a mixed list cost (a cost incurred by every stimulus on a mixed list, whether or not the task on that stimulus is different from the task on the stimulus before) and a shifting cost (a cost incurred only when the task on that stimulus is different from the task on the stimulus before). In addition, the mixed list cost is entirely in the baseline component, whereas the shifting cost has both a baseline component (the "tuning" cost) and an RSI component. Finally, the entire RSI component of the alternation cost appears to be a shifting cost. Experiment 4 and further analyses of Experiments 1 and 3 add more detail. In particular, task and response competition are part of the mixed list cost. Table 6 summarizes these findings.

In Experiments 5-8 we turned to understanding what occurs during the RSI that allows subjects to respond faster on alternating lists (but not as fast as on non-alternating lists). In short, we asked to what the RSI component corresponds. One basic distinction that was made was between a discrete switch and a gradual switch. A discrete switch holds that there is some switch process that occurs during the RSI that makes subjects more ready for one task than the other, and it must run its course before the selection of a response begins. Thus, RSI helps because the switch process can be completed during the RSI. A gradual switch, on the other hand, holds that subjects become more and more ready for the upcoming task over the RSI, and that once the stimulus is presented they begin selecting the response for the task. Thus, according to a gradual switch, RSI helps because subjects are more ready for the upcoming task the longer this gradual switch is allowed to operate.

In Experiment 5 subjects performed alternating and non-alternating task sequences on the first part of lists, but sometime between the 5th and 10th item the stimulus would be univalent (relevant to only one task) and might be for either task, regardless of the preceding sequence. When the instructed sequence was alternating but the task repeated on the the univalent stimulus, subjects were considerably slowed. This is consistent with the idea of a discrete switch, because with a discrete switch subjects are switching to the alternate task when they realize that the stimulus is for the task they just did. As a result they have to wait for this first switch to be complete and then switch back to the first task. It does not fit well, however, with a gradual set switch, since the subject in this case should be quite ready for the repeated task when the univalent stimulus is presented (at 0 ms RSI). It might be that the "surprise" of a repeated task when expecting an alternation of task slows subjects down. However, the gradual set switch cannot explain is that when a longer RSI is provided, the subjects are actually faster. A gradual set switch holds that during this RSI subjects are getting ready for the alternate task, so they should be even slower when the task repeats, even if there is a "surprise" factor that slows them down overall when the task unexpectedly repeats. A discrete switch has no problem explaining this because the RSI is used to perform the initial switch to the wrong but expected alternate task.

Another model of the RSI component that is consistent with this data that we have considered is a set decision model. According to this model the subject must make a decision as to which task he/she is going to do, and this decision occurs over the RSI. This model is very similar to a discrete switch model and, in fact, could even be considered one. We make the distinction for the following reason. A switch model invokes the notion of somehow tweaking some machinery to be better able to perform a task. In addition, it would suggest that every time one task is followed by another, the same switch should need to occur. The set decision model, however, can account for special situations in which the task alternates but the time cost associated with the RSI component is avoided.

Such a situation seems to occur in Experiment 6. In Experiment 6 subjects performed alternating and non-alternating sequences on the first 6 items in a list, and then responded to whichever task was relevant on the remaining 4 univalent items, in which the task for each of these items was randomly determined. We plotted RT to the final univalent item as a function of the task sequence on this item plus the previous three items (Figure 16B). This was called a set expectancy curve. It is clear from this data that subjects are using the preceding task sequence to predict (even though there is no predictive value in this) what the next task will be, and then they (somehow) commit to this task. Supposing, as seems natural, that the process that commits the subjects to one task or the other in this case is the same as the RSI component, the following problem arises. Subjects are presumably faster when the sequence is ABAB than when it is ABAA because they expect the task to alternate when the preceding sequence is ABA (i.e., the only difference between ABAB and ABAA is the task that occurs on the final trial). Thus, subjects tend to set themselves for the alternate task when the preceding sequence is ABA. Similarly, when the sequence is AAAA subjects are faster than when it is AAAB because when the preceding sequence is AAA they set themselves for a repetition of task. Thus, when the sequence is ABA* subjects perform an extra switch than when it is AAA* that, on average, buys them nothing. Thus, RT when the sequence is ABA* should be much slower (by the duration of the switch process) than when it is AAA*. However, it was actually slightly faster. This data, then, violates the view of a switch process that corresponds to the re-programming of a mechanism that selects responses. It is quite compatible, however, with the idea that a decision as to which task is to be performed must be made before the task is performed, and this decision sometimes does not take much time (as in the case where it is made depending on the previous task sequence).

These experiments lead to a view of the alternation cost in which various factors play a role. Task and response competition appear to play a role in explaining the mixed list cost. Criterion effects may also play a role here. In addition, there is slowing that is specifically due to the need to prepare the response selection machinery for one task or the other. These latter factors were embodied in the Ready, Set, Go! model (RSG model), which is re-stated here:

RSG1. There is a mechanism that selects responses, called the response selection mechanism, and this is the only way that responses are selected.

RSG2. This mechanism can be in a state that ranges from being not ready to do a particular task (and possibly ready to do a different one) to being fully ready for that task.

RSG3. The readiness for a particular task cannot be changed during "free" time (the RSI).

RSG4. The response selection mechanism is made ready for a task by that task being the last task performed (with the additional possibility that readiness slowly degrades to a neutral level with time).

RSG5. Before the response selection mechanism is used for a task, a decision as to which task it is to do must be made. This decision of which task to do -- call it setting the mechanism -- is not the same as being ready for a task. The response selection mechanism can be set for one task and ready (in the sense of RSG2-RSG4 above) for a different one.

RSG6. The setting of the response selection mechanism for one task or another is done during the RSI, and cannot be interrupted once started.

The single mapping hypothesis

In Section Three we tested the single mapping hypothesis. This hypothesis states that when learning a task subjects group S-R associations into separate mappings, and that the response selection mechanism can be set for only one mapping at a time. Thus, when subjects perform a different task from just before they will incur the switching costs described in the RSG model in the case where the two tasks are isolated in separate mappings, but not when all S-R associations for both tasks belong to the same mapping. In particular, this hypothesis predicts that the context in which two tasks are learned should have important influences on whether or not a cost of switching occurs when all stimuli on the lists are univalent.

This prediction was met in Experiments 9-12. Alternating list training resulted in no task repetition effect on random order lists (Experiment 11) and small or no alternation cost on alternating lists (Experiment 9 and 12). On the other hand, pure list training resulted in a 98 ms task repetition effect on random order lists (Experiment 10) and random order training similarly resulted in a 105 ms task repetition effect (Experiment 10) and an 82 ms alternation cost (Experiment 12). These data are well explained by the single mapping hypothesis if one supposes that when subjects learn two tasks on alternating lists they learn to group the tasks into a single mapping, but when they learn two tasks on pure lists or random order lists they learn to group each task into a different mapping.

There is another reason that the single mapping hypothesis makes sense. Suppose that it was not true, and subjects could set themselves for an arbitrary collection of S-R associations. Then, they could set themselves for naming the color of a stimulus if a color is presented and subtracting three from a number if a number was presented. If subjects are then presented with and respond to the color, the RSG model would hold that the response selection mechanism now becomes more "tuned" for both the color and subtraction task. Although not impossible, this to us seems like a very odd situation: Given time to prepare for the subtraction task, the response selection mechanism cannot be "tuned". But having thought that a moment ago one might have had to subtract three from a number, but instead having had to name a color, does better "tune" the mechanism for this task.

Concluding remarks

We do not want to claim that we have solved the problem of task set with the RSG model and the single mapping hypothesis. However, we do believe that we have accurately described at a very broad level the mechanisms involved. In addition, the level of description that we have provided seems well suited for further flushing out using computational techniques. For example, it is not obvious how task competition, response competition and "tuning" effects can all be attributed to the same mechanism. Perhaps they are not. This is not inconsistent with the RSG model. However, it seems likely to us that they are. In this case, understanding what types of mechanisms can exhibit these effects (additively) would perhaps put greater limits on the detailed workings of this machinery.

Still, one limitation of this work, in particular, should be acknowledged. This limitation is that all the experiments involved arbitrary 4-choice manual RT tasks. It certainly is not outrageous to assume that the findings with these tasks would transfer to other stimulus and response modalities and to non-arbitrary tasks. Generalizations originally derived with manual choice-RT responses in the PRP paradigm, for example, have transferred to (at least punctate) verbal naming tasks. But it is also possible that certain features of our conclusions are particular to the tasks we used. For example, it was discussed in Section Three that the single mapping hypothesis may not apply when the tasks are non-arbitrary. The suggestion is that non-arbitrary tasks can be ad hoc added to the current task set. One thing that would seem to follow from this is that there should be no baseline component to the alternation cost, i.e. the alternation cost at long RSI should vanish.

It is too early to tell whether the above modification to the model is needed. If it is, it should be emphasized, it would not take the wind out of the sails of the model. Indeed, it would be a welcome and fascinating amendment.

Figure 1: Two models of processing when alternating tasks. SN denotes when stimulus N is made; RN denotes when response N is made. Time flows from left to right. (A) Processing is shown for non-alternating task lists. The hypothetical A, B, and C stages each begin when the previous one is finished. (B) The full discrete switch model: processing on B stages are postponed until a switch process is complete. Each successive switch begins at the beginning of the previous B stage. Responses are made at a slower rate because of the inserted switch process. (C) The constant readiness model: Each stage begins when the previous one ends as on non-alternating lists. One or more stages are longer in duration than on non-alternating lists (only stage 2 here), accounting for why responses are made at a slower rate.

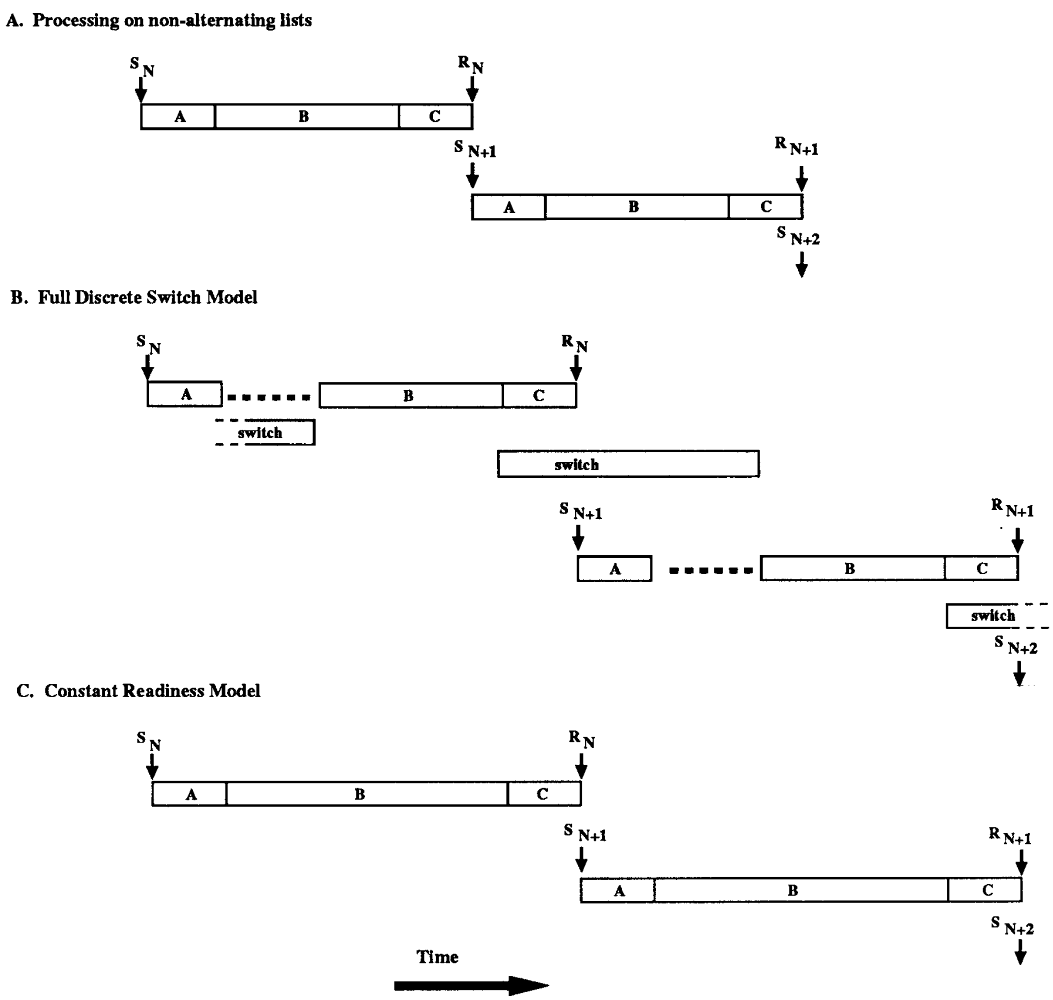

Figure 2: Mean correct RT in Experiment 1 broken down by display condition (0 or 1500 ms RSI or preview) and alternation condition.

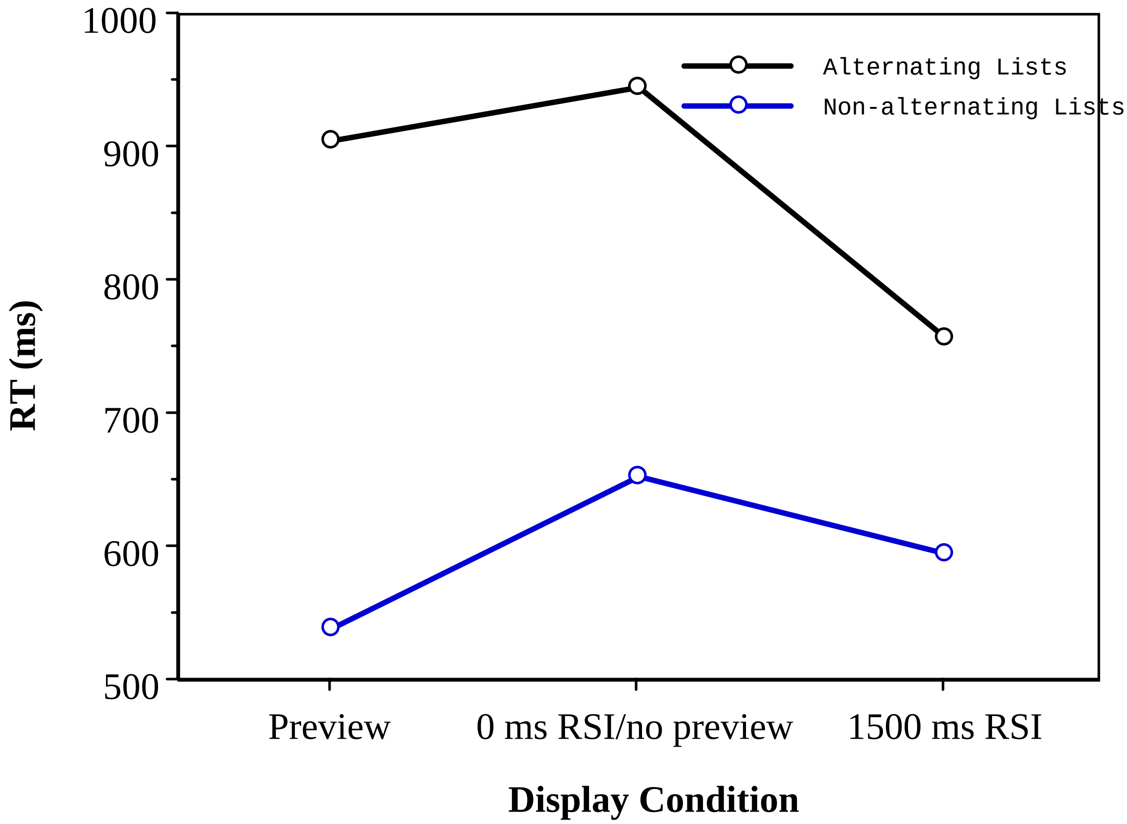

Figure 3: RT distributions for alternating and non-alternating lists in the 0 ms RSI condition of Experiment 1.

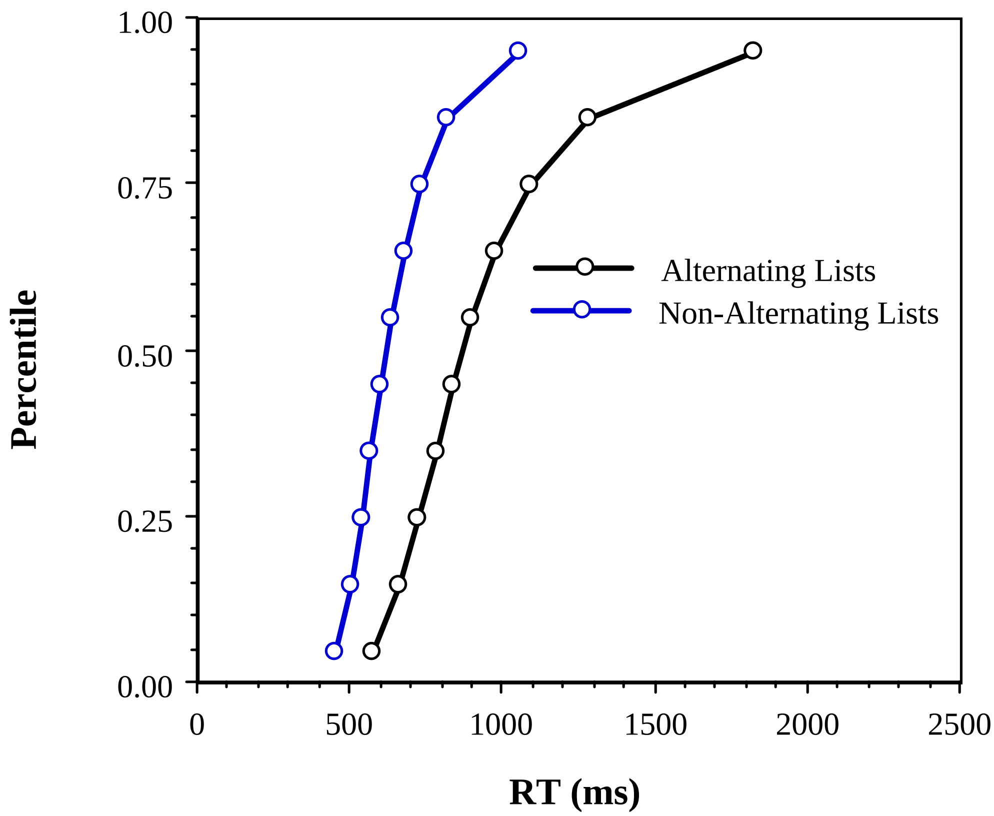

Figure 4: Theoretical readiness functions for four different models. (A) The full discrete switch model. (B) The constant readiness model. (C) The partial discrete switch model. (D) The gradual shift model.

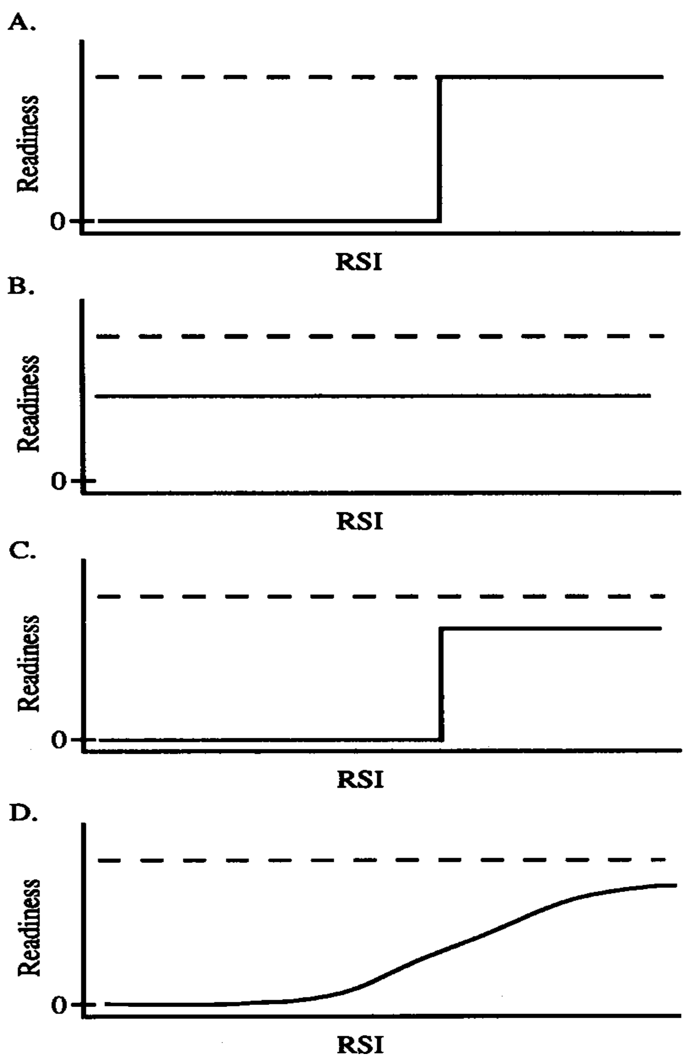

Figure 5: Mean correct RT for alternating and non-alternating lists in Experiment 2 as a function of RSI. Open symbols are for when RSI varied within the list; filled symbols are for when RSI varied between lists.

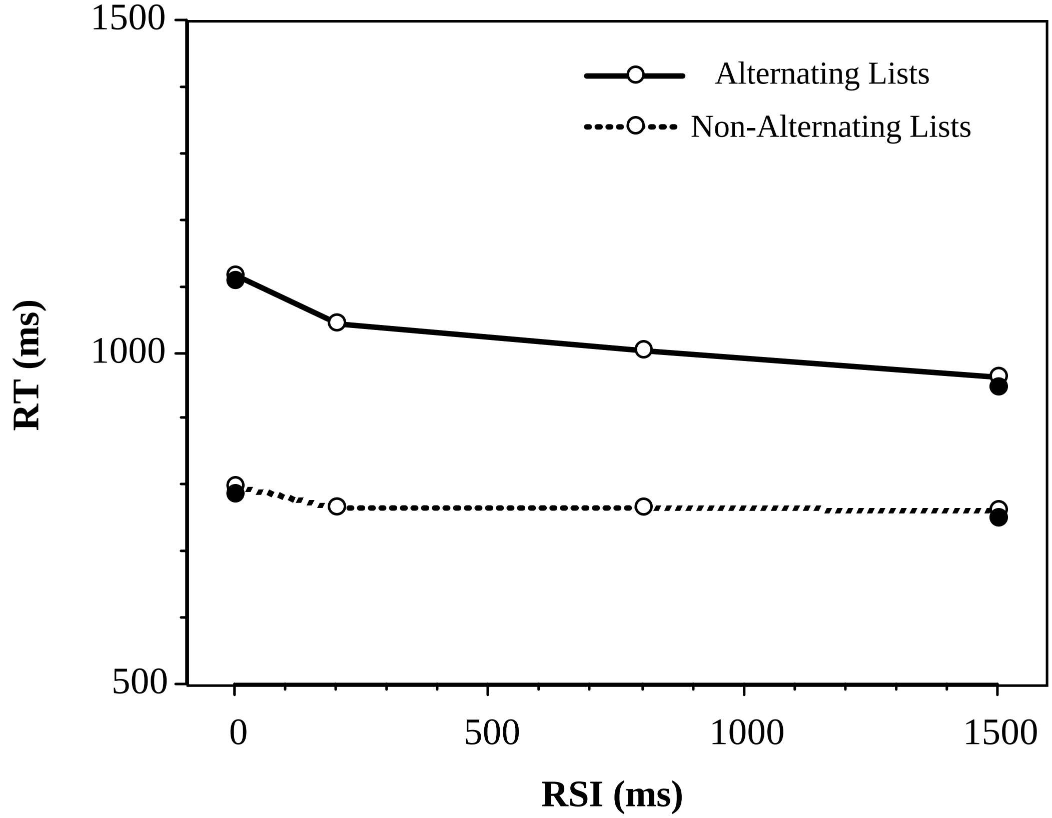

Figure 6: Ratio of readiness to asymptotic readiness as a function of RSI in Experiment 2 derived by assuming asymptotic readiness is reached by the time a response is made on each trial.

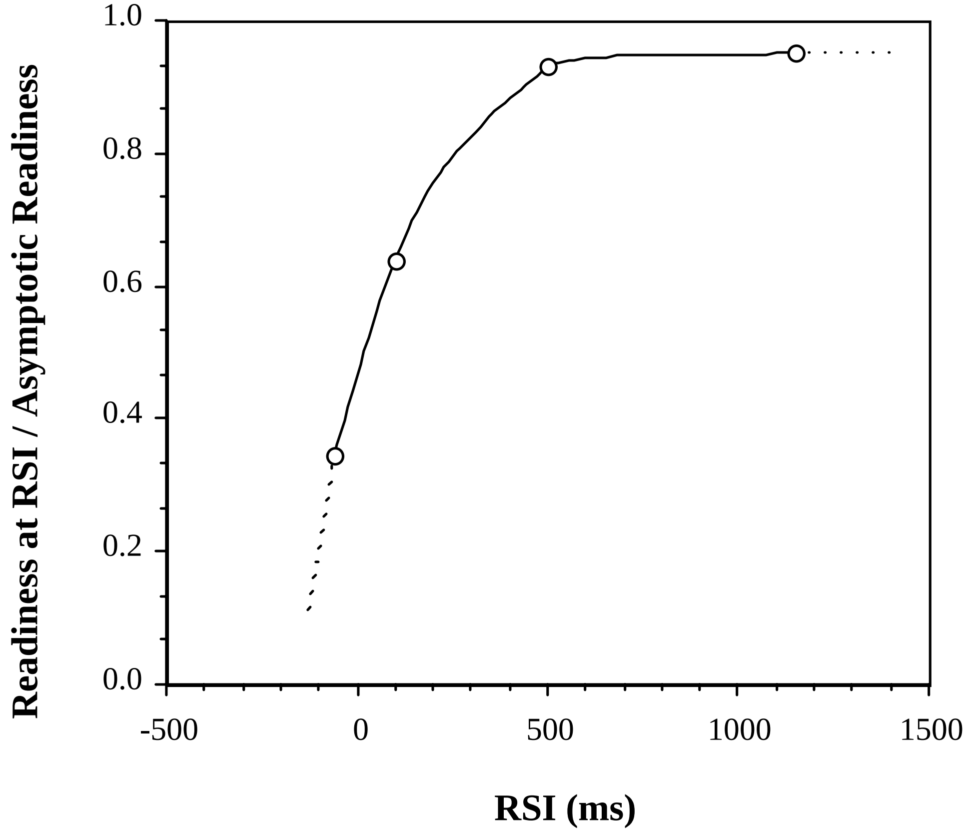

Figure 7: Mean correct RT to AAAA, ABAB, AABB-same, and AABB-different stimuli in Experiment 3 as a function of RSI.

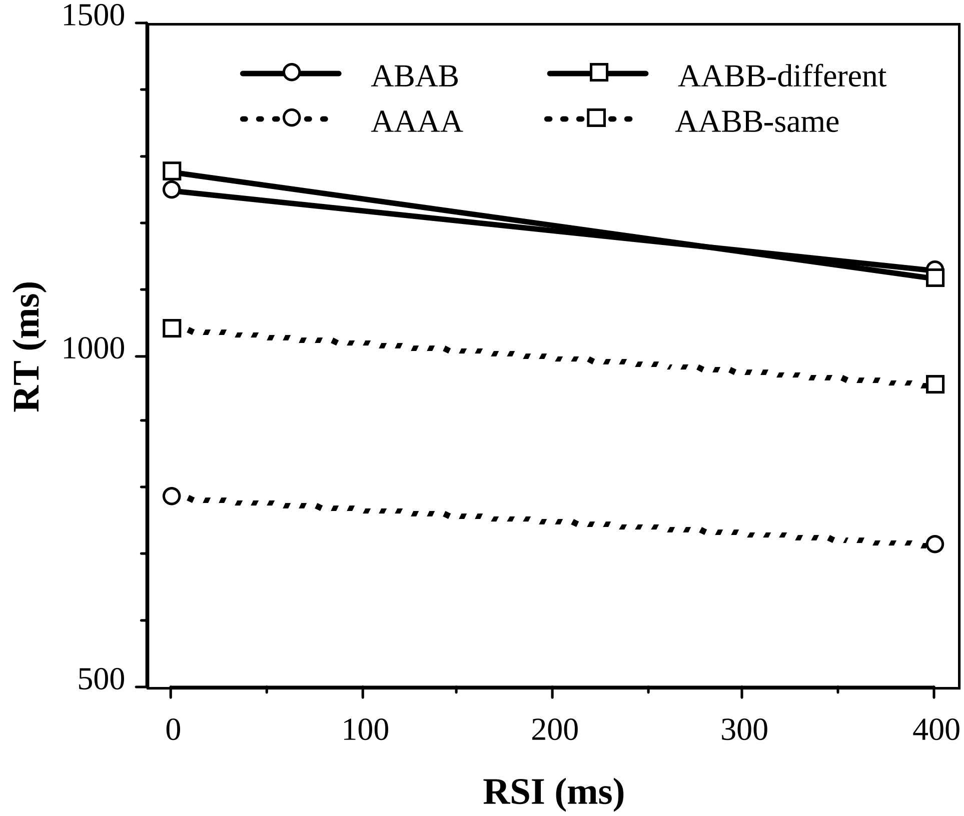

Figure 8: Mean correct RT in Experiment 1 (A) and 3 (B) as a function of sequence, compatibility, and RSI.

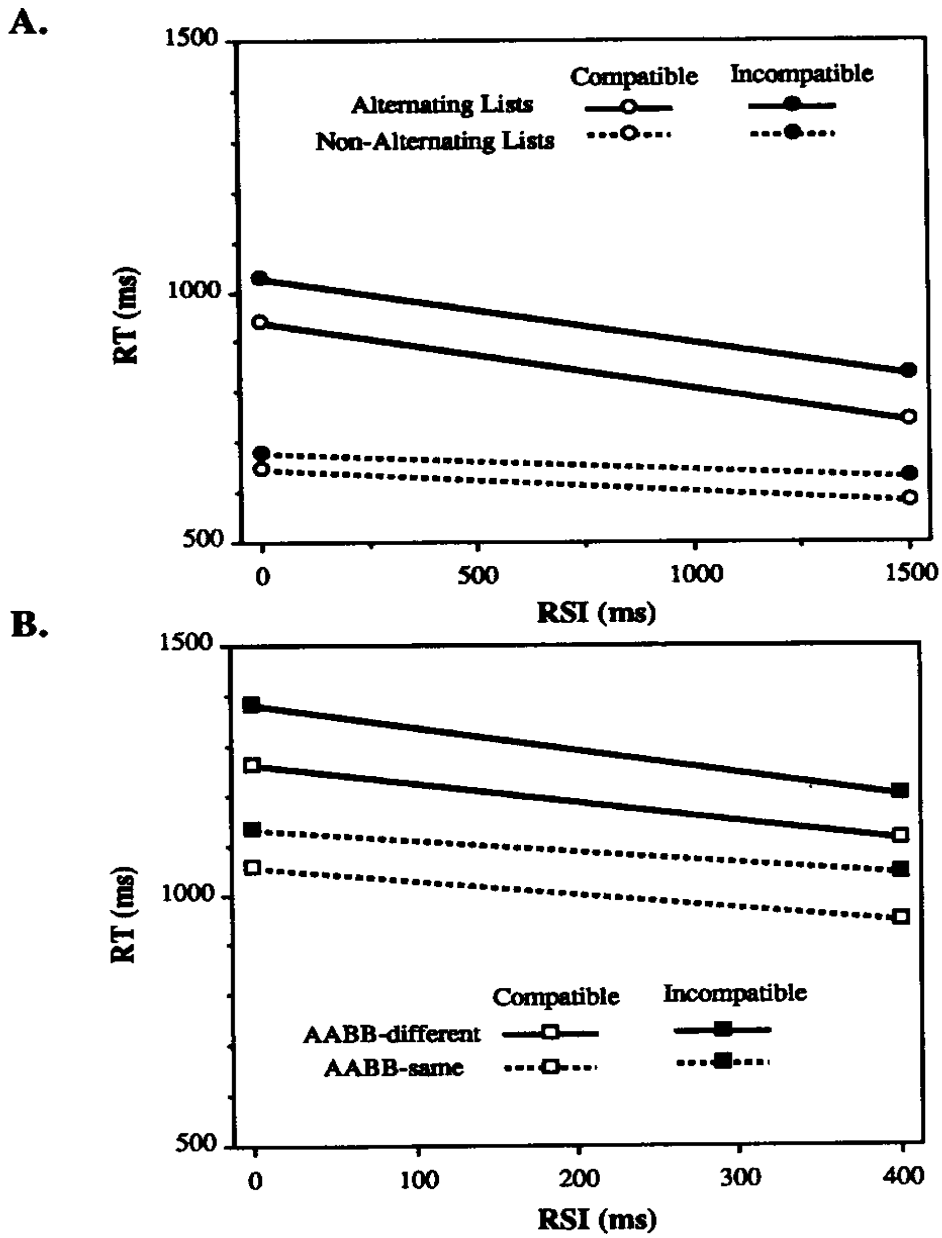

Figure 9: Mean correct RT on alternating and non-alternating lists in Experiment 4 as a function of stimulus type (compatible, incompatible, univalent).

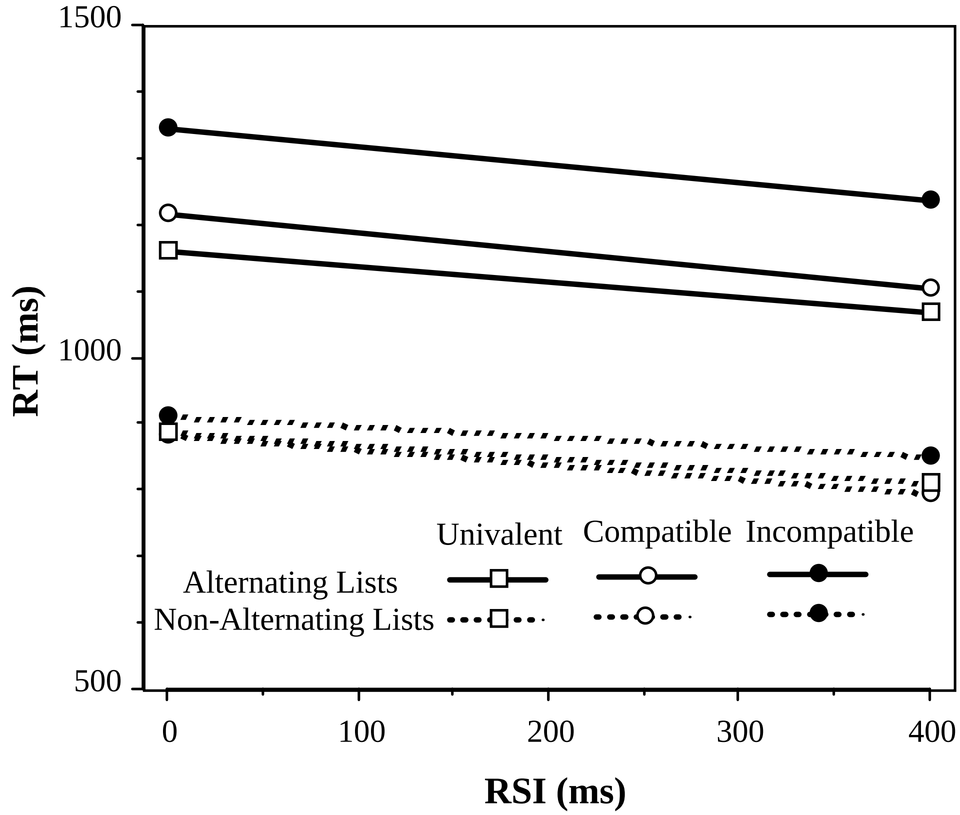

Figure 10: The interaction with RSI of various factors tested in Experiments 1,3, and 4. Response competition is based on an average of the effects from Experiments 1,3, and 4, weighted by number of subjects.

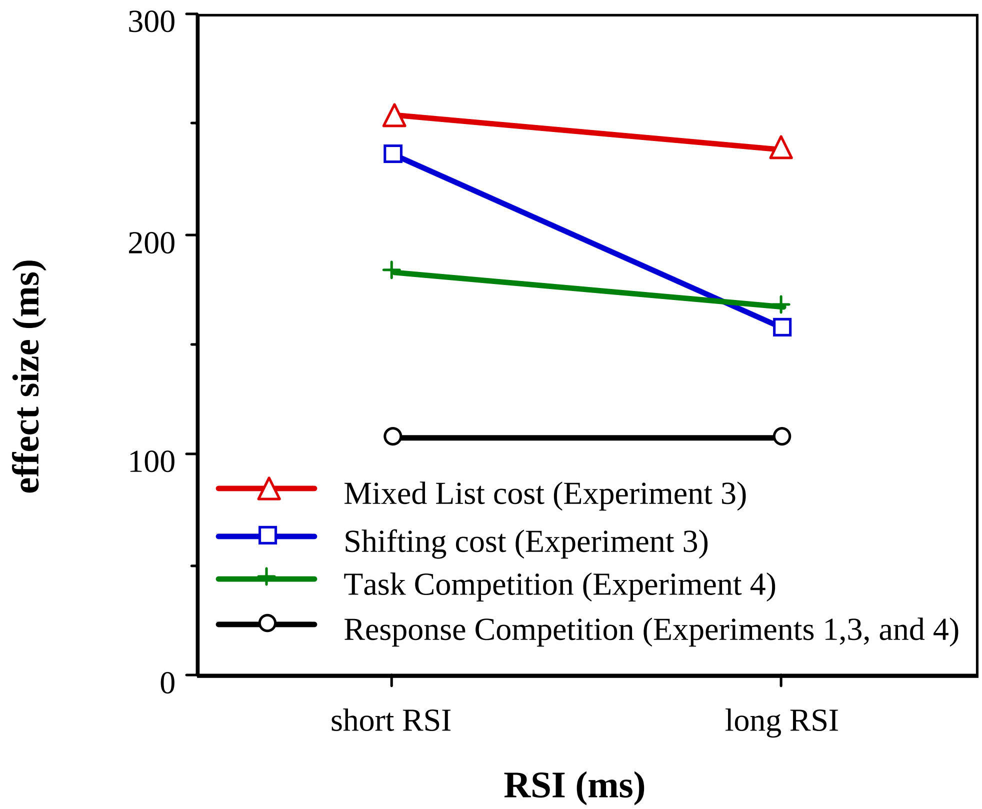

Figure 11: The baseline component of the alternation cost broken down into the different factors that account for it. The shifting cost is estimated from Experiment 3; response and task competition is estimated from Experiment 4; the left-over RSI component is estimated from Experiment 2.

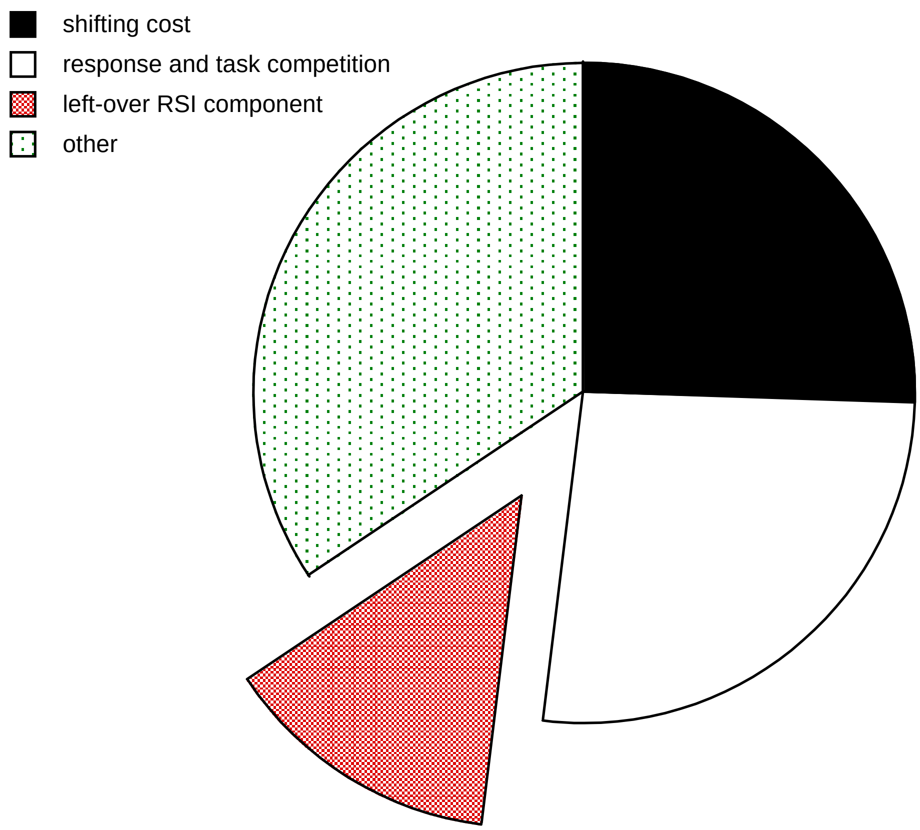

Figure 12: Mean correct RT on the bivalent part of the list in Experiment 5 as a function of task sequence and RSI.

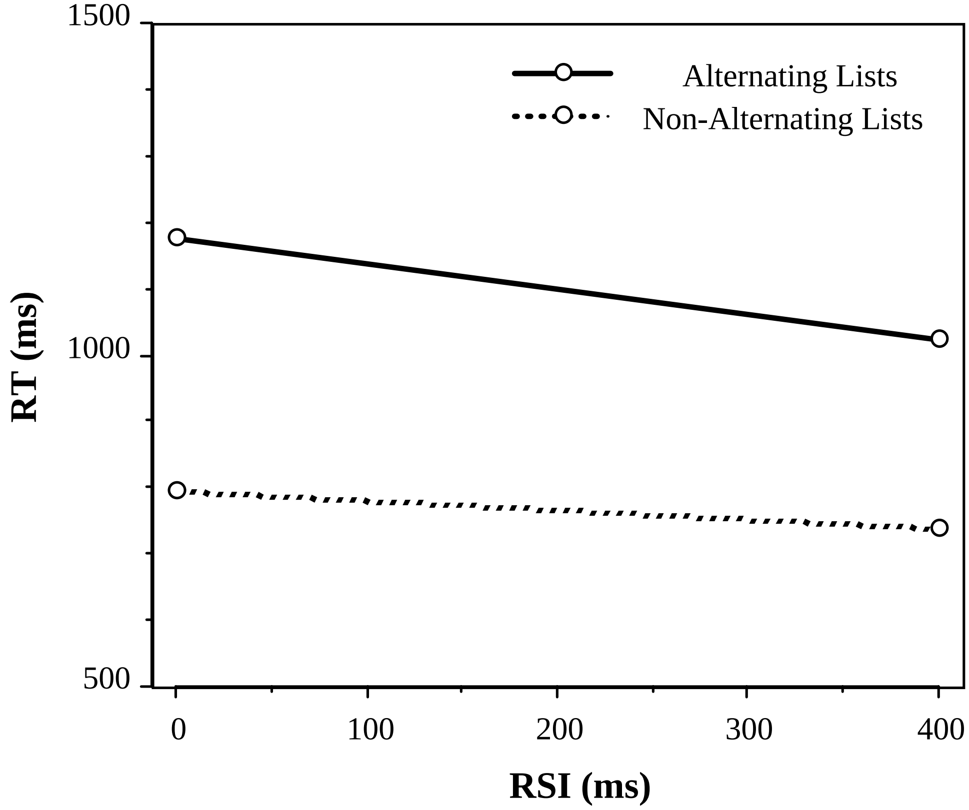

Figure 13: Mean correct RT to univalent stimuli in Experiment 5 as a function of the bivalent task sequence and whether the previous task (last bivalent task) was the same or different from the univalent task (RSI = 0 ms only).

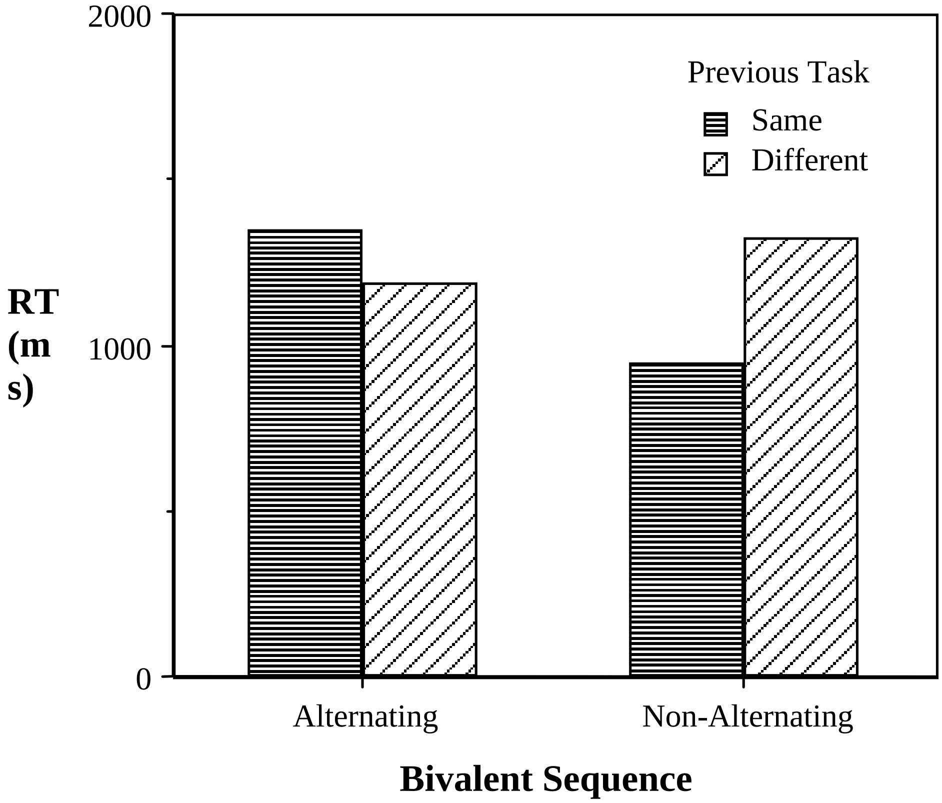

Figure 14: Mean cost/benefit of RSI as a function of bivalent task sequence and whether the previous task (last bivalent task) was the same or different from the univalent task. (Cost of RSI is negative, benefit is positive).

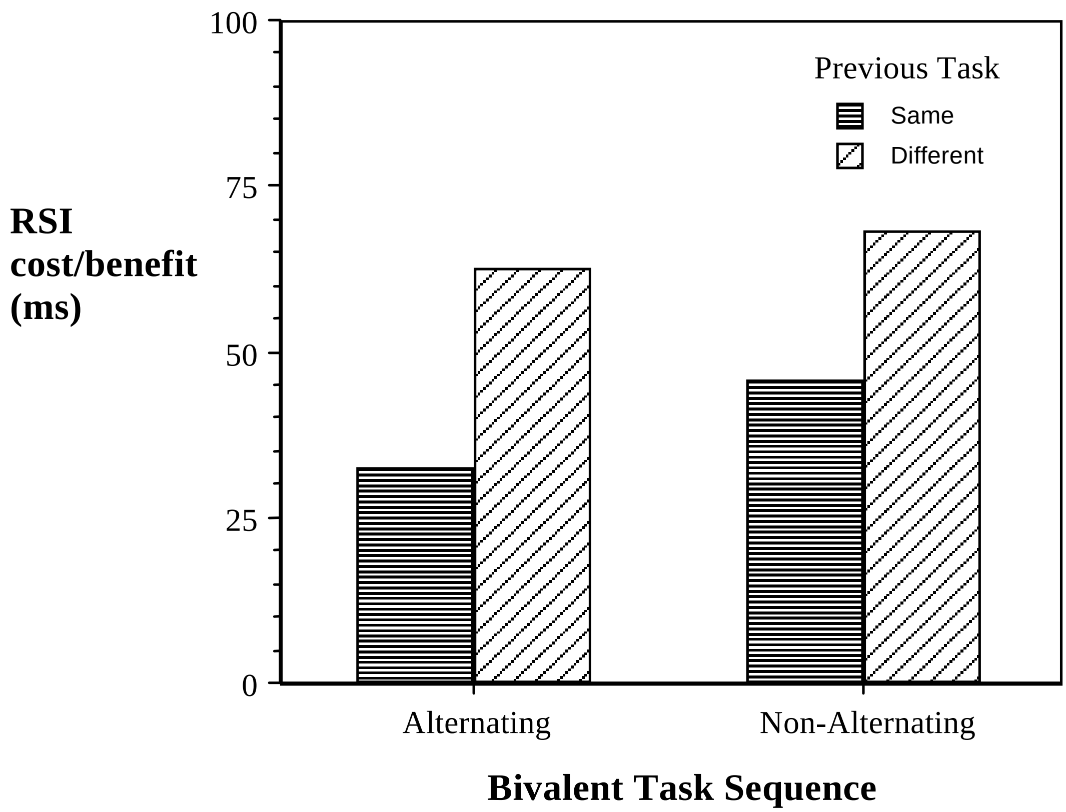

Figure 15: Mean correct RT to univalent stimuli for alternating and non-alternating bivalent task sequences in Experiment 6 as a function of response number. (A) Only those trials where the univalent task sequence plus the last bivalent task happen to be non-alternating. (B) Only those trials where the univalent task sequence plus the last bivalent task happen to be alternating. (C) As a function of whether or not the previous task was same or different from the present one.

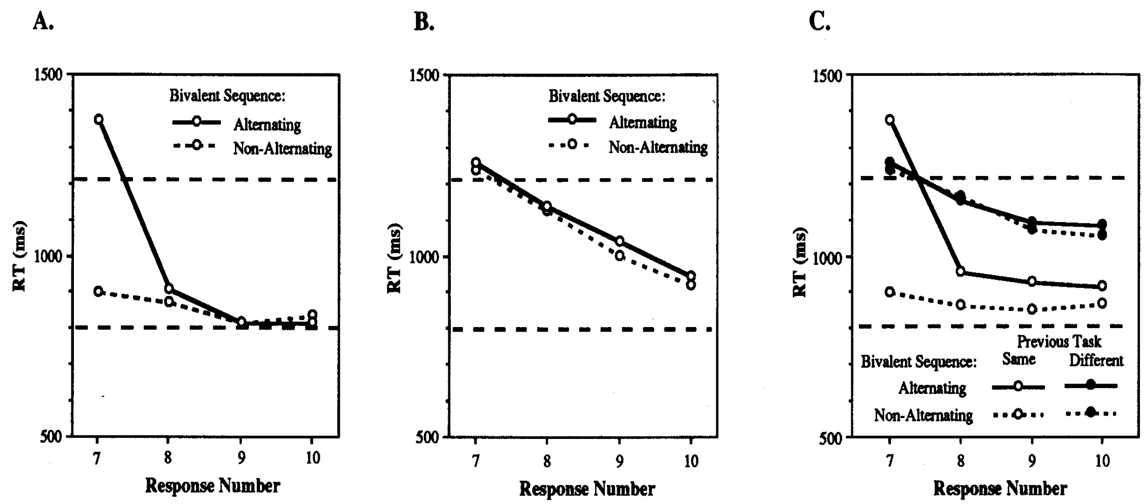

Figure 16: Expectancy curves: (A) Stimulus expectancy curve: RT in a serial 2-choice RT task for short RSI (50 ms after key up) and long RSI (50 ms after key down) as a function of preceding stimulus sequence (data from Vervaeck and Boer, 1980, Experiments 1 and 2). (B) Set expectancy curve: RT to the final univalent stimulus on each list in Experiment 6 as a function of the preceding task sequence. On the ordinates of the figure "A" stands for the stimulus/task 3 back and "B" stands for the other stimulus/task.

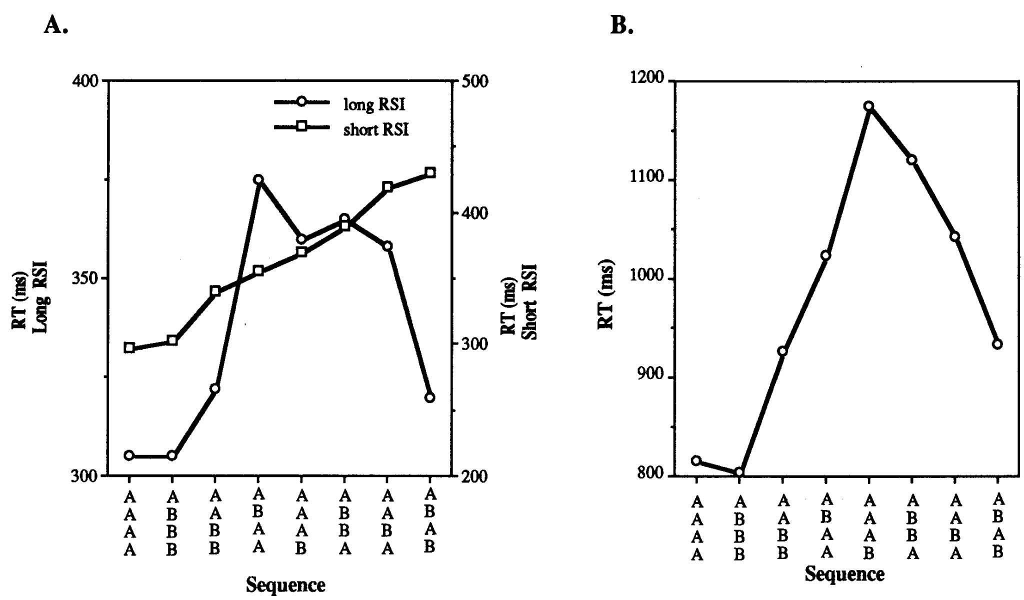

Figure 17: Mean correct RT for alternating and non-alternating bivalent task sequences in Experiment 7 as a function of stimulus type (bivalent stimulus, univalent stimulus for the same task as the stimulus before, univalent stimulus for the other task from the stimulus before).

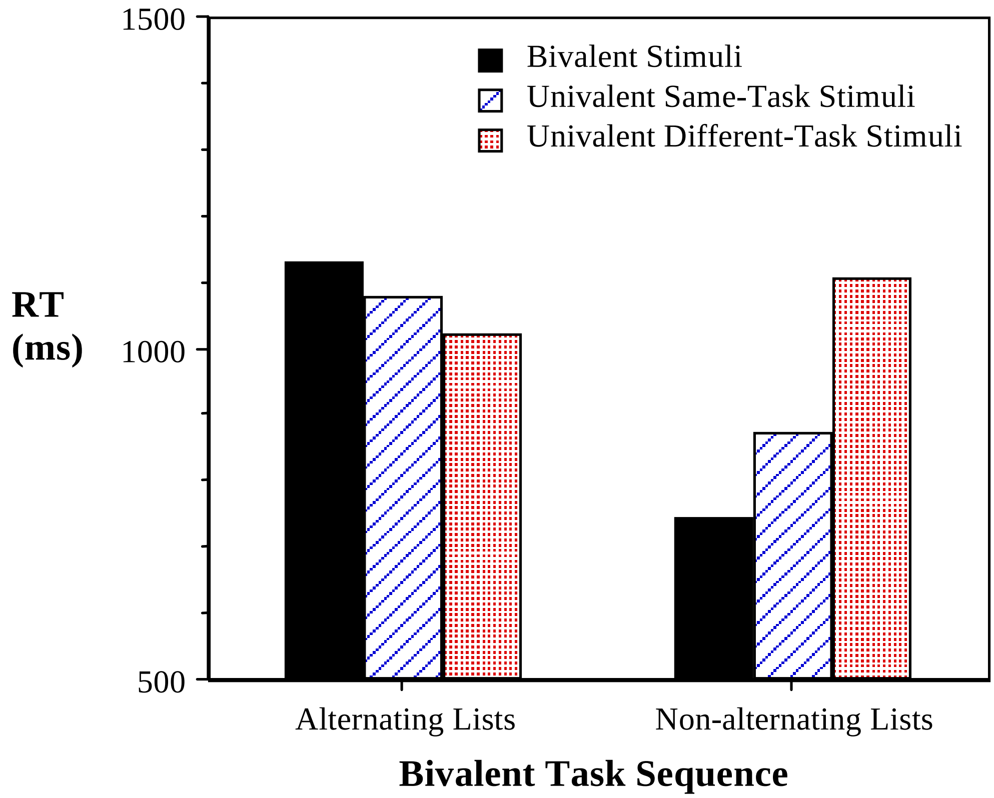

Figure 18: Mean correct RT (A) and error rates (B) in Experiment 8 as a function of whether the sequence is fixed or random, whether the previous task is the same or different from the present one, and SOA.

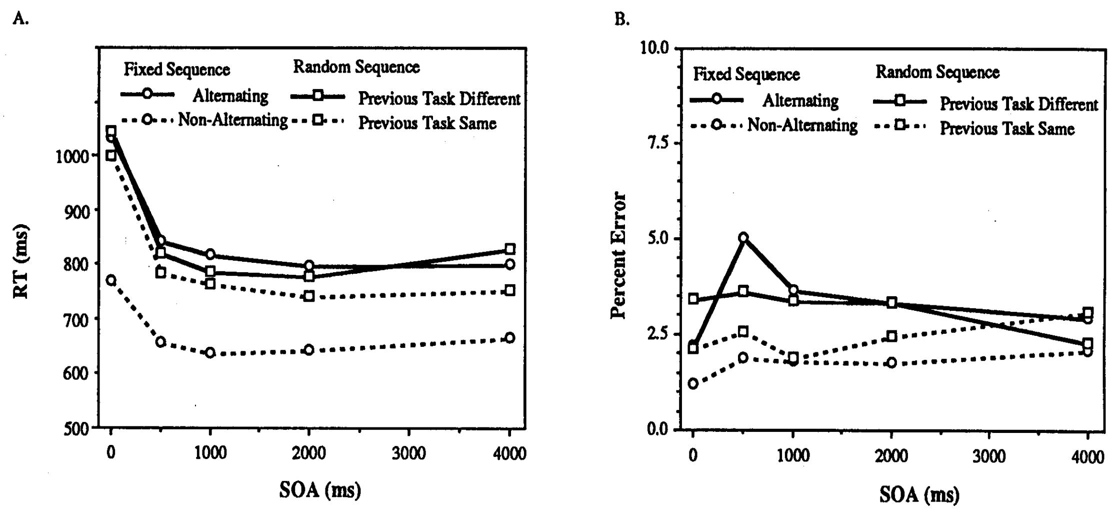

Figure 19: Mean correct RT for pure training and mixed training groups in Experiment 10 as a function of block number and, during the test phase (blocks 15-24), whether the previous task was the same or different.

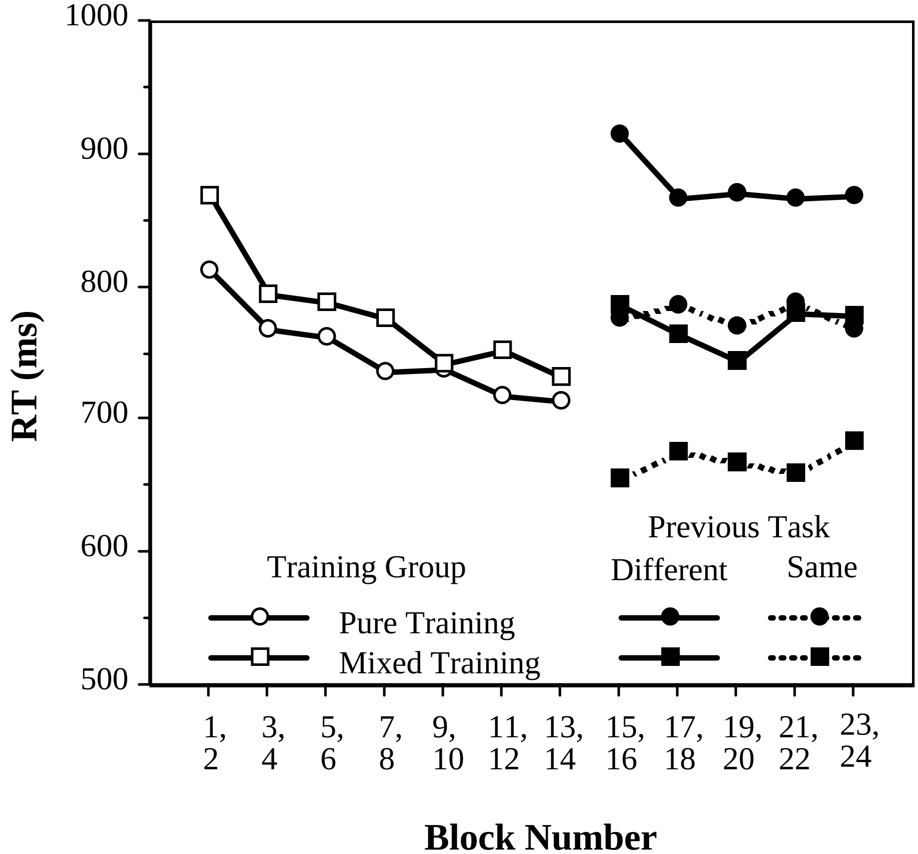

Figure 20: Expectancy curves: (A) Stimulus expectancy curve: RT in a serial 2-choice RT task for short RSI (50 ms after key up) and long RSI (50 ms after key down) as a function of preceding stimulus sequence (data from Vervaeck and Boer, 1980, Experiments 1 and 2). (B) Set expectancy curve: RT to the final univalent stimulus on each list in Experiment 6 as a function of the preceding task sequence. On the ordinates of the figure "A" stands for the stimulus/task 3 back and "B" stands for the other stimulus/task.

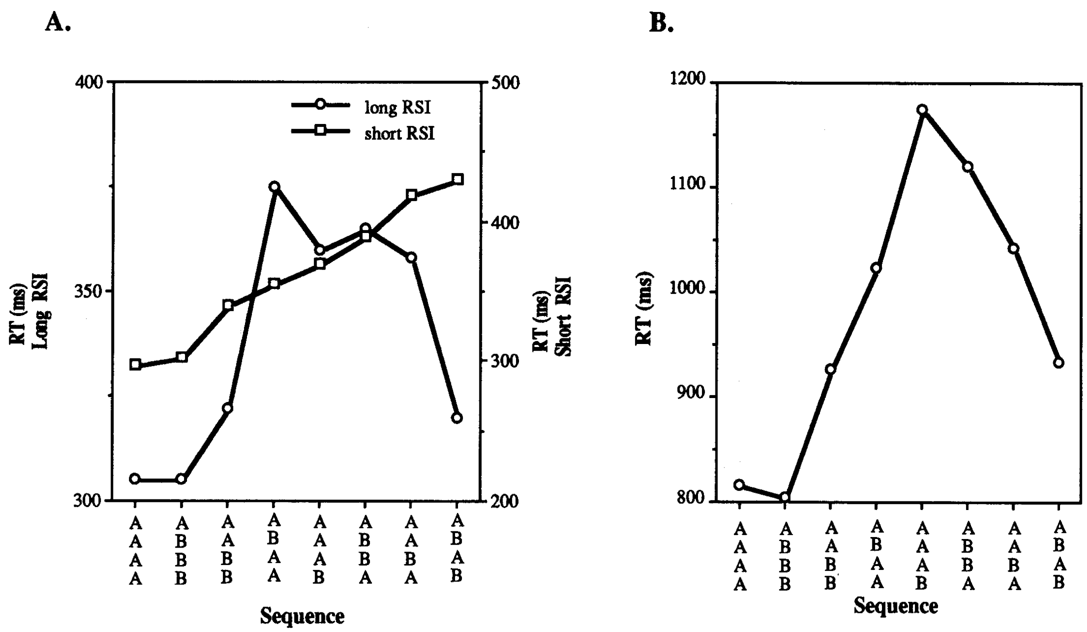

Figure 21: Set expectancy curves for pure task and mixed task training in Experiment 10.

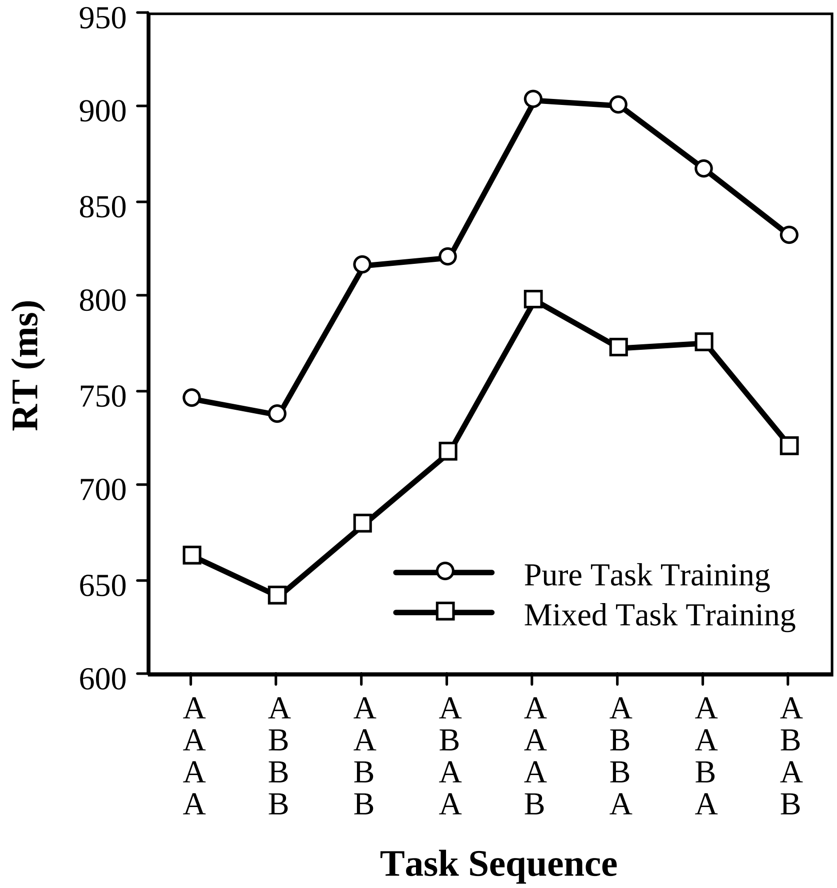

Figure 22: Mean correct RT for alternating task training in Experiment 11 as a function of block number and, during the test phase (blocks 15-24), whether the previous task was the same or different.

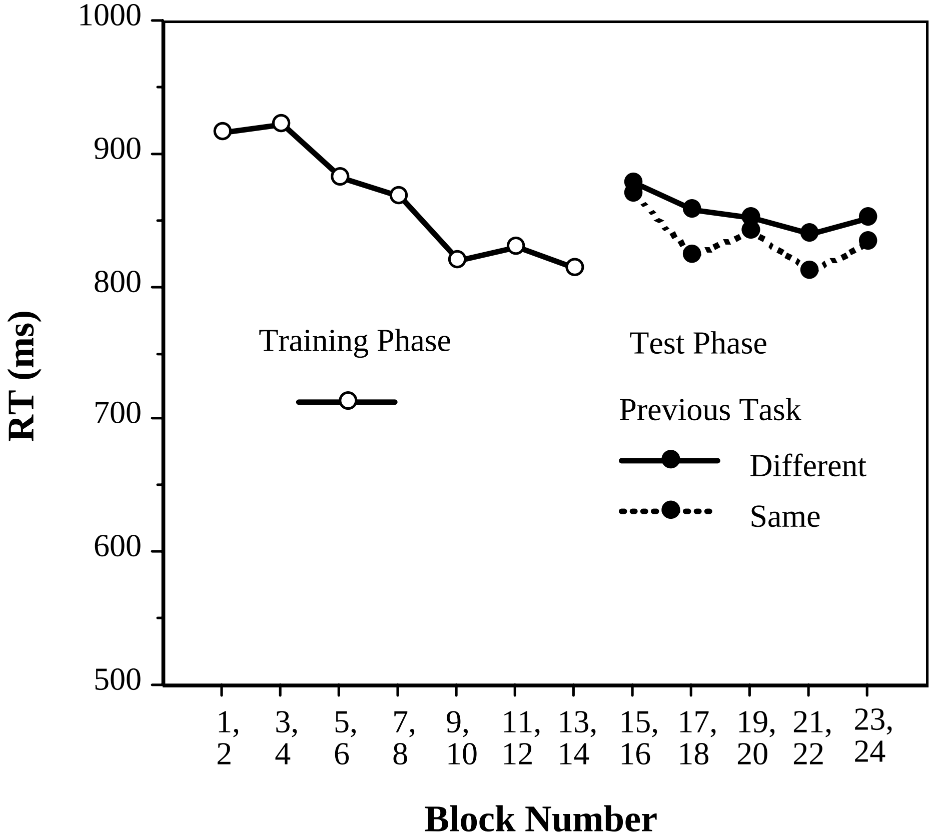

Figure 23: Set expectancy curves for alternating task training in Experiment 11.

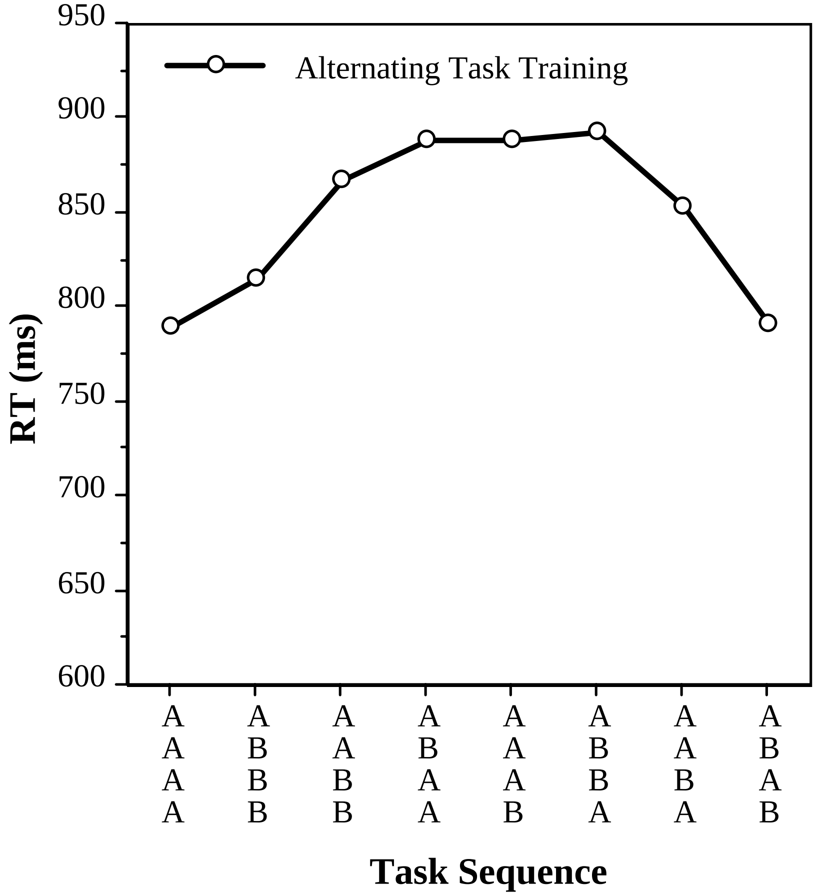

Figure 24: Mean correct RT during the training phase for random order, AABB list, and alternating list training groups in Experiment 12 as a function of block number and pure versus mixed lists.

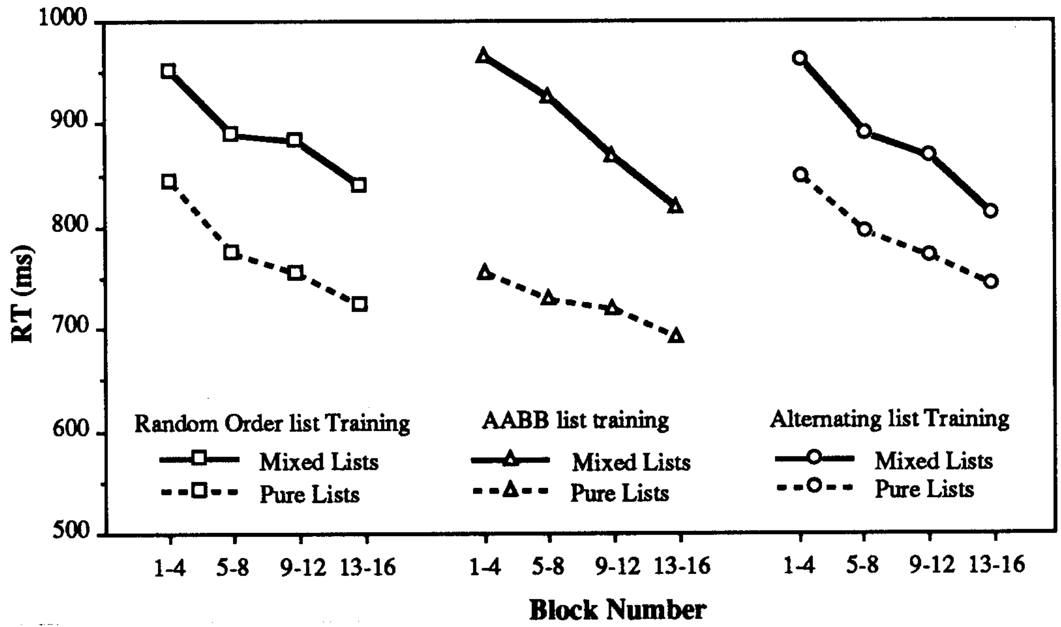

Figure 25: Mean correct RT on mixed lists during the training phase for random order and AABB list training groups in Experiment 12 as a function of block number and same versus different previous task.

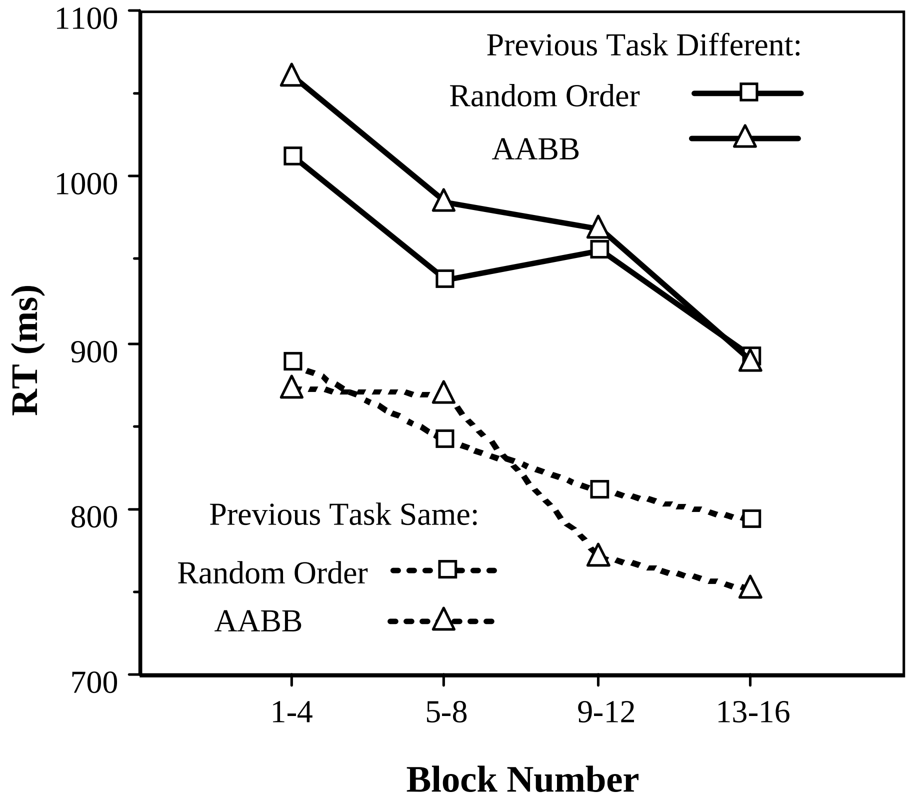

Figure 26: Mean correct RT during the test phase for random order, AABB list, and alternating list training groups in Experiment 12 as a function of block number and alternation (mixed lists for each training group are alternating lists during the test phase).

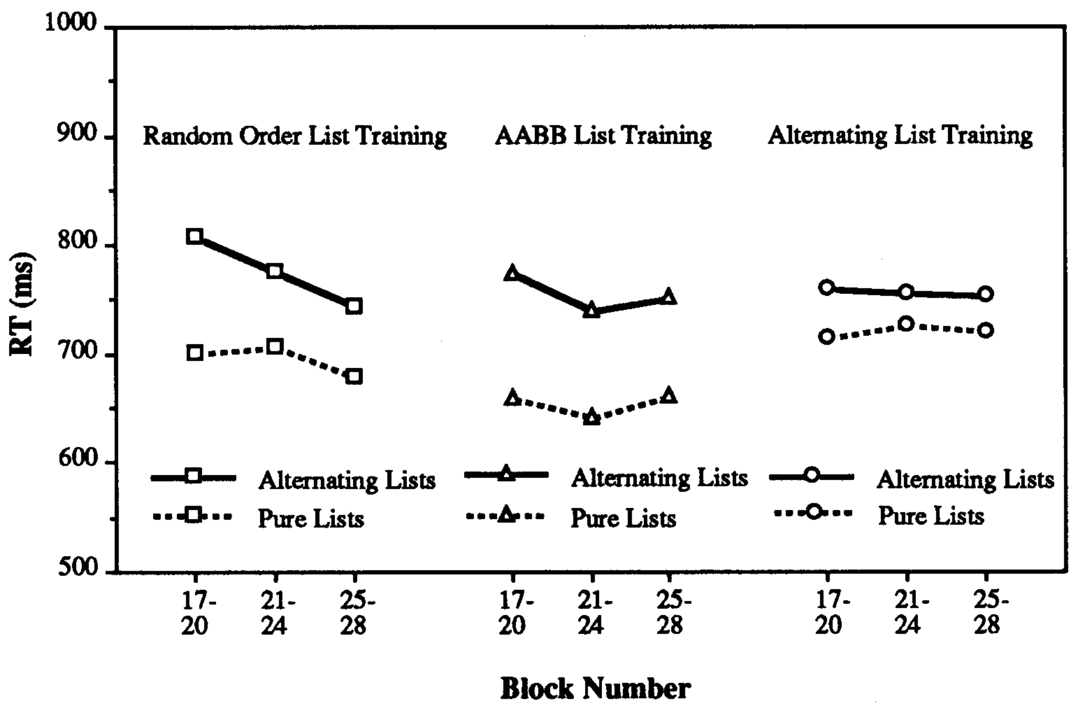

Table 1: Mean number of times a trial had to be restarted in Experiment 1 as a function of display condition and sequence.

	Sequence	
Display Condition	Alternating	Non-alternating
Preview	.152	.168
0 ms RSI/no preview	.245	.172
1500 ms RSI	.196	.190

Table 2: Estimated error rate per stimulus (in percent) in Experiment 1 as a function of display condition and sequence.

	Sequence	
Display Condition	Alternating	Non-alternating
Preview	1.4	1.5
0 ms RSI/no preview	2.0	1.5
1500 ms RSI	1.6	1.6

Table 3: Summary of various terms used in the text with their defining formulas and a brief description of to what they refer.

Term	Formula	Description
Alternation cost	ABAB - AAAA	total cost associated with alternating sequence
Baseline component	ABAB - AAAA at long RSI	part of alternation cost that cannot be overcome with time
RSI component	alternation cost at short RSI - alternation cost at long RSI	part of alternation cost that can be overcome with time
shifting cost	AABB-different - AABB-same	cost associated with doing a different task from just before
mixed list cost	AABB-same - AAAA	cost associated with doing a task on a list that has more than one task on it (both task A and task B)
bivalent stimuli	----	stimuli relevant to both tasks
univalent stimuli	----	stimuli relevant to only one task
task competition	bivalent RT - univalent RT	cost associated with having two activated tasks relevant to current stimulus
response competition	RT to response incompatible stimuli - RT to response compatible stimuli	specific form of task competition that only occurs when tasks lead to different responses on the current stimulus

Table 4: Break-down of the alternation cost into RSI and baseline component and mixed list and shifting cost, and which causes might affect which of the four parts.

	RSI	Baseline
mixed list cost	set decision response & task competition criterion effect	response & task competition criterion effect
shifting cost	set switch set decision response & task competition	"tuning" cost response & task competition

Table 5: Estimated by-item error rate per stimulus (in percent) in Experiment 3 as a function of key condition and sequence. (Average number of restarts per list are in parentheses).

	Sequence		
Key Condition	AAAA	ABAB	AABB
Same Keys	1.8 (0.21)	3.9 (0.50)	2.9 (0.37)
Different Keys	1.4 (0.15)	1.8 (0.20)	2.2 (0.26)

Table 6: Break-down of the alternation cost into RSI and baseline component and mixed list and shifting cost, and which causes might affect which of the four parts after Experiments 3 and 4.

	RSI	Baseline
mixed list cost	-------------	response & task competition criterion effect
shifting cost	set switch set decision	"tuning" cost task competition

Table 7: Mean number of times a trial had to be restarted in Experiment 5 as a function of task sequence, RSI, and key condition.

		Sequence	
Key Condition		Alternating	Non-alternating
Same Keys	0 ms RSI	0.45	0.22
	400 ms RSI	0.41	0.31
Different Keys	0 ms RSI	0.31	0.18
	400 ms RSI	0.21	0.16

Table 8: Estimated error rate per stimulus (in percent) in Experiment 6 as a function of task sequence and valence. (Average number of restarts per list, broken down by whether the restart occurred on a univalent or bivalent stimulus, are shown in parentheses).

	Sequence	
Valence	Alternating	Non-alternating
Univalent	2.4 (0.10)	2.0 (0.09)
Bivalent	3.9 (0.30)	2.4 (0.17)

Table 9: Estimated error rate per stimulus (in percent) in Experiment 7 as a function of task sequence and valence. (Average number of restarts per list, broken down by whether the restart occurred on a univalent or bivalent stimulus, are shown in parentheses).

	Sequence	
Valence	Alternating	Non-alternating
Univalent	4.2 (0.045)	4.5 (0.048)
Bivalent	3.7 (0.18)	2.2 (0.10)

References

Allport, A. & Styles, E.A. (1991). Multiple executive functions, multiple resources? Experiments in shifting attentional control of tasks. Unpublished manuscript, Oxford University.

Cattel, J.M. (1886). The time it takes to see and name objects. Mind, 11, 63-65.

Duncan, J. (1977). Response selection rules in spatial choice reaction tasks. In Dornic, S. (Ed.), Attention and Performance VI. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Forrin, B. (1974). Naming latencies to mixed sequences of letters digits. In P. M. A. Rabbitt (Ed.), Attention and Performance V. London: Academic Press.

Hick, W. E. (1952). On the rate of gain of information. Quarterly Journal of Experimental Psychology, 4, 11-26.

Hyman, R. (1953). Stimulus information as a determinant of choice reaction time. Journal of Experimental Psychology, 45, 188-196.

Jersild, A.T. (1927). Mental set and shift. Archives of Psycholgy whole no. 89.

Kafrey, D., & Kahneman, D. (1977). Capacity sharing and refractoriness in successive reactions. Perceptual and Motor Skill, 44, 327-335.

Krueger, L. E., & Shapiro, R. G. (1981). Intertrial effects of same-different judgements. Quarterly Journal of Experimental Psychology, 33A, 241-265.

Leonard, J.A. (1953). Advance information in sensori-motor skills. Quarterly Journal of Experimental Psychology, 5, 141-149.

Marcel, T., & Forrin, B. (1974). Naming latency and the repetition of stimulus categories. Journal of Experimental Psychology, 103, 450-460.

Pashler, H. (1994). Overlapping mental operations in serial performance with preview. Quarterly Journal of Experimental Psychology, 47A, 161-191.

Pashler, H., & Baylis, G. (1991A). Procedural learning: 1. Locus of practice effects in speeded choice tasks. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 20-32.

Pashler, H., & Baylis, G. (1991B). Procedural learning: 2. Intertrial repetition effects in speeded choice tasks. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 33-48.

Rabbitt, P.M.A. (1969). Psychological refractory delay and response-stimulus interval duration in serial, choice-response tasks. Acta Psychologica, 30, 195-219.

Rabbit, P.M.A. (1980). The effects of R-S interval duration on serial choice reaction time: Preparation time or response monitoring time? Ergonomics, 23, 65-77.

Rabbit, P.M.A., & Yvas, S. (1973). What is repeated in the "repetition effect"? In S. Kornblum (Ed.), Attention and Performance IV. New York: Academic Press.

Schmidt, R.A., & Bjork, R.A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. Psychological Science, 3, 207-217.

Spector, A. & Biederman, I.(1976). Mental set and mental shift revisited. American Journal of Psychology, 89, 669-679.

Stroop, J. R. (1935). Studies of interference in serial-verbal reaction. Journal of Experimental Psychology, 18, 643-662.

Sudeven, P., & Taylor, D. (1987). The cuing and priming of cognitive operations. Journal of Experimental Psychology: Human Perception and Performance, 13, 89-103.

Vervaeck, K.R., & Boer, L.C. (1980). Sequential effects in two-choice reaction time: Subjective expectancy and automatic after-effect at short response-stimulus intervals. Acta Psychological, 44, 175-190.

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.

Welford, A. T. (1959). Evidence of a single-channel decision mechanism limiting performance in a serial reaction task. Quarterly Journal of Experimental Psychology, 11, 193-210.

Wilkinson, R.T. (1990). Response-Stimulus interval in choice serial reaction time: Interaction with sleep deprivation, choice, and practice. Quarterly Journal of Experimental Psychology, 42A, 401-423.

Appendix: Computing the by-item error rates based on the number of re-starts per list

In this Appendix we describe how the by-item error rates can be computed using the trial re-start data. The basic approach is to express the expected number of trial re-starts per list as a function of a fixed by-item error rate. This function is then inverted in order to yield by-item error rate as a function of expected number of trial re-starts. By-item error rates are then estimated with this equation for each subject by substituting the average number of re-starts per list for the expected number of re-starts per list.

We begin by deriving the formula for computing the by-item error rates. After this we will address questions of statistical bias. Finally, we will discuss how to compute separate by-item error rates for the beginning of a list vs. the end of a list. This is used in Experiments 6 and 7 in order to derive separate by-item error estimates for bivalent and univalent stimuli.

By-item error rates as a function of the expected number of re-starts per trial

Let p_1 be the by-item error rate and p_N be the chance of making an error somewhere on a list of N items. These are related by

(1) $$p_N = 1 - (1 - p_1)^N.$$

The list will have to be re-started at least once with probability pN. In particular, the list will be re-start exactly once with probability p_N * $(1-p_N)$, exactly twice with probability p_N * $(1-p_N)$^2, and exactly k times with probability p_N * $(1-p_N)$^k. Thus, the expected number of re-starts per list is:

(2) $$ER = p_N \sum_{k=1}^{\infty} k(1 - p_N)^K = \frac{1 - p_N}{p_N}$$

Combining 1 and 2 and solving for p_1 yields

(3) $$p_1 = 1 - \sqrt[N]{\frac{1}{1 + ER}}$$

Statistical bias in the estimate of by-item error rates

The by-item error rate can be estimated with (3) by substituting the average number of re-starts per list for the expected number of re-starts per list. This is not, however, an unbiased statistic. We performed several simulations in order to assess the amount of statistical bias. In each of these simulations the statistic was applied on data from runs of 10 lists of 10 items (the number of lists and items per list in each condition of Experiment 1). Then, 30,000 runs were executed, and the data from these runs averaged. The standard error of the mean was negligible in all cases, and thus the numbers listed below can be taken as accurate well beyond the number of digits reported.

First, we performed the simulation with the assumption that the error rate is the same throughout the experiment. In two different runs we used error rates of 1.5% and 2.5%. These numbers seemed representative of our data. The simulations yielded estimates of the by-item error rates of 1.45% and 2.41%, respectively. Thus, there is a minor underestimate of the error rates using the by-item error statistic.

It might be, however, that the error rate varies within the list. How will this affect our estimates? To test this we performed a simulation in which the error rate on a given item was randomly chosen from 0.5%, 1.5%, and 2.5% with equal probability. The actual error rate (the error rate on lists that were not restarted) under these circumstances was 1.49%. It is not an average of the three error rates above because a re-start is more likely when the error rate is higher, so lists with all low error rates are more likely to be completed without a re-start. The by-item error statistic was 1.44%. Thus, when the error rate varies within the list, there is also a small bias towards under-estimating the error rate.

Finally, error rate might vary between lists. Thus, some lists might be performed very sloppily, others very conservatively. We ran two simulations under these conditions. In the first simulation the error rates were chosen from a small range (0.5%, 1.5%, or 2.5%); in the second simulation the error rates were chosen from a large range (1.0% or 10.0%). The actual and estimated error rates in the small range simulation were 1.43% and 1.41 %, respectively. Thus, though there is still a small bias in this simulation, it is reduced. The actual and estimated error rates in the large range simulation were 3.49% and 4.4%, respectively. Thus, the statistic <u>over</u>-estimates the actual error rate here.

In summary, if the error rate varies within each list as much as it varies between lists, there is a minor underestimate of error rate with the error statistic derived from (3). If error rate is fixed within lists but varies a little between lists, then there is also a slight, but now smaller, bias towards underestimating error rate. However, if the error rate is fixed within lists and varies substantially between lists, the statistic overestimates the error rates.

The biases that are present, then, are not a problem if one wants to put an upper-bound on error rate. That is, the conditions under which the error rate is underestimated do not result in large a underestimation. In fact, the bias is removed in most cases when the numbers are rounded to the nearest 0.1%. However, if error rates do vary between but not within lists, the derived error rates may not be good estimates of the actual error rates -- they may over-estimate the error rate substantially.

Separate estimates of by-item error rate for the beginning and end of a list

It may sometimes be desirable to separately estimate the chance of making an error on the first k items in a list and the remaining N-k items. For example, in

Experiment 6 the first 6 items were bivalent and the last 4 were univalent; in Experiment 7 the first 4 items were bivalent and the last item was univalent. We wanted to compare the chance of an error on a univalent item to the chance of an error on a bivalent item.

We can estimating these two quantities using the average number of re-starts per list that occur on the first part of the list versus the second part of the list. We will call these quantities RA and RB, respectively. The by-item error rate for the last part of the list can be computed as in (3) with RB substituted for ER. This is because a re-start that occurs on the first part of the list will not effect the number of re-starts that occurs on the last part of the list. The by-item error rate for the last part of the list can then be estimated with:

$$pB1 = 1 - \sqrt[N-k]{\frac{1}{1+RB}}$$

Computing the error rate for the first part of the list is a bit trickier. This is because a re-start on the last part of the list may effect the number of re-starts on the first part of the list, i.e. subjects must start over at the beginning of the list even if the error is on the last part of the list. The quantity that we then want to plug into (3) is the average number of times the list had to be re-started before the first part of the list was successfully completed just once. Call this RA*. Let pB be the chance of making an error somewhere on the last part of the list (one time through). Then:

$$ERA^* = ERA + pB \cdot ERA + pB^2 \cdot ERA + \ldots = ERA \sum_{k=0}^{\infty} pB^k .$$

Based on the formula for the sum of a geometric progression,

$$ERA^* = \frac{ERA}{1 - pB}.$$

So,

$$ERA = (1 - pB) \cdot ERA^*.$$

Finally, the by-item error rate for the first part of the list can be estimated with:

$$pA1 = 1 - \sqrt[k]{\frac{1}{1+\frac{RA}{1+RB}}}$$

Experiment to Set Mapping:

Experiment 1	Set1
Experiment 2	Set17
Experiment 3	Set25
Experiment 4	Set18B
Experiment 5	Set21
Experiment 6	Set22
Experiment 7	Set23
Experiment 8	Set27
Experiment 9	Set12
Experiment 10	Set24
Experiment 11	Set24B
Experiment 12	Set28