Pashler, H. (2000). Task switching and multitask performance. To appear in Monsell, S., and Driver, J. (editors). Attention and Performance XVIII: Control of mental processes. Cambridge, MA: MIT Press.

12

Task Switching and Multitask Performance

Harold Pashler

ABSTRACT Research on task switching and dual-task performance has spawned two literatures that have, to a surprising extent, developed independently. This tutorial reviews the principal findings of each tradition and considers how these phenomena may be related. Beginning with Jersild 1927, task-switching studies reveal that when people perform two tasks in succession, with each task requiring different responses to the same set of stimuli, substantial slowing occurs. Recent research suggests that while this slowing can be partially ameliorated by allowing sufficient time between tasks, advance reconfiguration is almost always incomplete. In studies of dual-task performance, stimuli arc presented very close together in time, and subjects attempt concurrently to perform two wholly distinct tasks. A substantial slowing of one or both tasks is usually observed. The most stubborn source of this slowing appears to be queuing of central processing stages, sometimes supplemented by other kinds of interference. This queuing occurs even when the tasks are highly dissimilar and is unlikely to reflect voluntary strategies. A number of possibilities for how task switching and dual-task queuing plight be related are discussed critically, including the possibility that queuing might stem from an inability to maintain two distinct task sets at the same time.

What happens when people try to switch rapidly between one task and another? What happens when they try to do more than one task at the same time? The first of these two fundamental questions is chiefly discussed in a modest-sized literature under the label "task switching" or "mental set"; the second, in a much larger literature under the label "divided attention" or "dual-task performance." The present chapter reviews main phenomena and theoretical issues in both areas and tries to draw some substantive connections between them.

12.1 TASK SWITCHING

In 1927, well before the modern era of information-processing psychology, an educational psychologist named Arthur T. Jersild published a pioneering study of people's ability to alternate between different tasks. Jersild measured the total time it took a person to work through a printed list of stimuli, making a response of some kind to each individual item in turn. In pure task blocks, subjects performed the same task on each item (for example, subtracting three from each number on a list). In alternating-task blocks, subjects performed one task on all the oddnumbered stimuli, and another task on the even-numbered stimuli. In some of the experiments, every stimulus was a potential input for either task (following Fagot 1994, I will refer to this arrangement of tasks and stimuli as a "bivalent" list or mapping). One of Jersild's bivalent alternating-task lists contained two-digit numbers; subjects were instructed to subtract three from the first number, add six to the second number, subtract three from the third, and so forth. They were substantially slower (more than 0.5 sec per item) in bivalent alternating lists than in pure lists, sometimes by more than 1 sec per item. This difference between pure and alternating bivalent lists will be referred to as the "alternation cost."

Jersild also examined the case of task alternation, where each stimulus was a potential input only for the appropriate task (henceforth referred to as a "univalent" list or mapping). For example, one univalent alternating list contained two-digit numbers and words, numbers alternating with words; subjects were instructed to subtract three from each number and to say aloud the antonym of each word. Remarkably, Jersild found that there was no alternation cost at all with these univalent lists; indeed, subjects were actually slightly *faster* in responding to alternating lists than to pure lists.

Some fifty years later, Spector and Biederman (1976) confirmed Jersild's basic results, finding a modest-sized benefit of alternation with univalent lists. This occurred, however, only when the items were printed as in Jersild's studies, and subjects were allowed to preview items ahead of the ones they were responding to.[1] When the items were placed on cards, so that subjects could not see the next stimulus until they turned a card over, there was actually a small alternation cost; the same was true when the experiment was run with a discrete-trials procedure. With the alternating bivalent lists (adding three, then subtracting three, etc.), Spector and Biederman found a large alternation cost (402 msec/item). This was cut about in half, to 188 msec/item, when a visual task cue ("+3" or "-3") was placed next to each item.

Several rather trivial potential explanations for the basic alternation cost need to be considered.

One might propose that the alternation cost merely reflects a tendency of subjects occasionally to forget what task they should perform next. If this is correct, the slowing should largely be confined to a few, very slow trials. This does not seem to be the case, however. Fagot (1994) had subjects make button-push responses to either the color or the identity of a letter (an *A*, *B*, or *C* in red, green, or blue). Figure 12.1 shows the Vincentized reaction time (RT) distributions for a zero response-stimulus interval (RSI) condition where the two tasks were performed in alternation.[2] The slowing is by no means confined to the slowest responses. Evidently, then, among the sources of the alternation cost are events that occur on at least a significant number of trials.

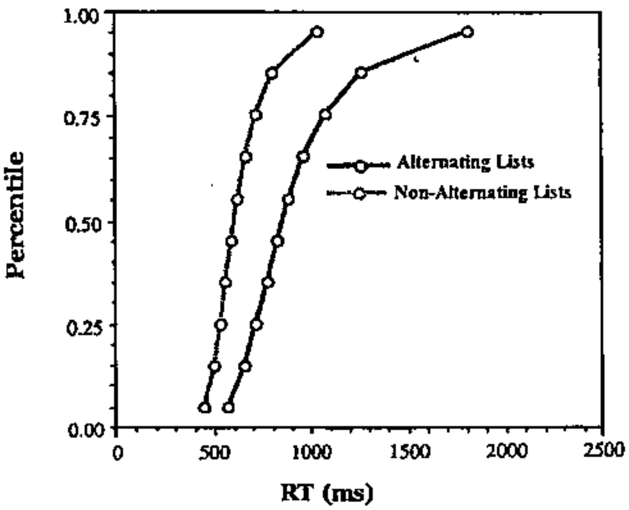

Figure 12.1 Vincentized reaction time (RT) distributions for a bivalent list alternating-task design. Alternation cost appears even among the fastest responses. From Fagot 1994; reprinted with permission.

One might also propose that the faster responses to pure as opposed to alternating lists arise because alternating lists do not include any stimulus repetitions. In any speeded-choice task, people respond much faster to stimuli that match whatever was presented on the preceding trial (Kornblum 1973; Pashler and Baylis 1991). This potential confound does not explain the effect, however. In the experiment by Fagot (1994) shown in figure 12.1, lists were selected with the constraint that there be no item repetitions, but the alternation cost was still found; the same was probably done informally in some of the earlier studies.[3]

What, then, accounts for the alternation cost with bivalent lists, and why is this cost sometimes virtually absent with univalent lists? Perhaps the most obvious interpretation is that depicted in figure 12.2. According to this "task set reconfiguration" (TSR) view (Monsell 1996), preparing to perform a task involves linking and/or configuring different processing modules. Different modules are assumed to be responsible for different aspects or stages of the task (e.g., perception, response selection, etc.). With bivalent lists, task alternation requires changing the links, settings, or both between when the central processing of one stimulus is completed and when the central processing of the next begins. In some cases, changes in the configuration of perceptual modules may also be involved. Given the conflicting response selection rules in the case of bivalent stimuli, the module responsible for response selection cannot be set the same way throughout the block of trials. At first blush in this account, one would assume that the alternation cost simply reflects the time needed to complete the switch. As for univalent alternating tasks, it should be possible for the two task mappings to coexist more or less

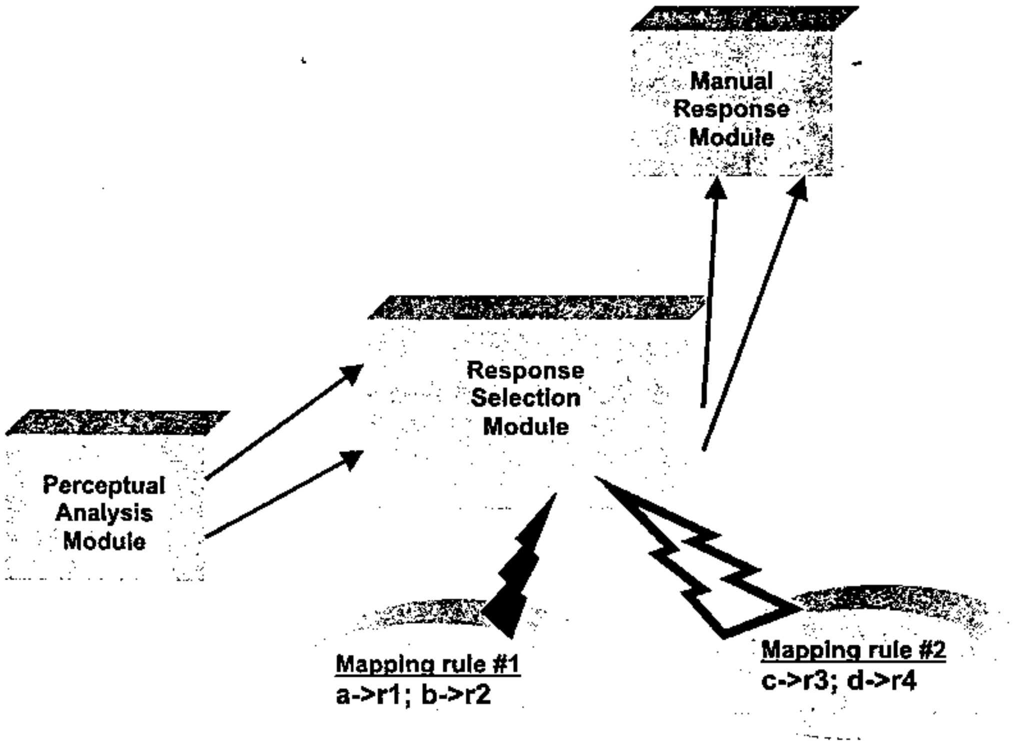

Figure 12.2 Discrete conception of task set switching. The response selection machinery is prepared at any one time to perform either of the two incompatible mappings, but not both. In the alternating-task blocks, one mapping is switched out and the other inserted, somewhat as a crystal in early radio sets.

happily, so that the union of the two mappings could simply be loaded into the response selection module. This may explain why there should be minimal cost in that situation, although of course by itself it does not explain why there should ever be a benefit.

If this account is correct, allowing extra time between the response to stimulus *n* and the presentation of stimulus *n* + 1 (RSI) might allow subjects to complete the switch in advance, thereby reducing or eliminating the alternation cost. Many recent studies of task alternation have found some reduction. A notable example is Rogers and Monsell 1995, which found an approximately 50% reduction as RSI was lengthened from 150 msec to 1,200 msec, so long as subjects could rely on having the long RSI. In Fagot 1994; conducted in my own laboratory, subjects were instructed • to respond to colored letters by pushing a button to indicate either the color of the letter or its identity. In alternating-task blocks, RSI varied from 0 to 1.5 sec; the alternation cost fell from 314 msec at the zero RSI to 201 msec at the 1.5 sec RSI, with most of the reduction occurring over the range of RSIs between zero and 400 msec; this pattern was confirmed in several other experiments within that series.

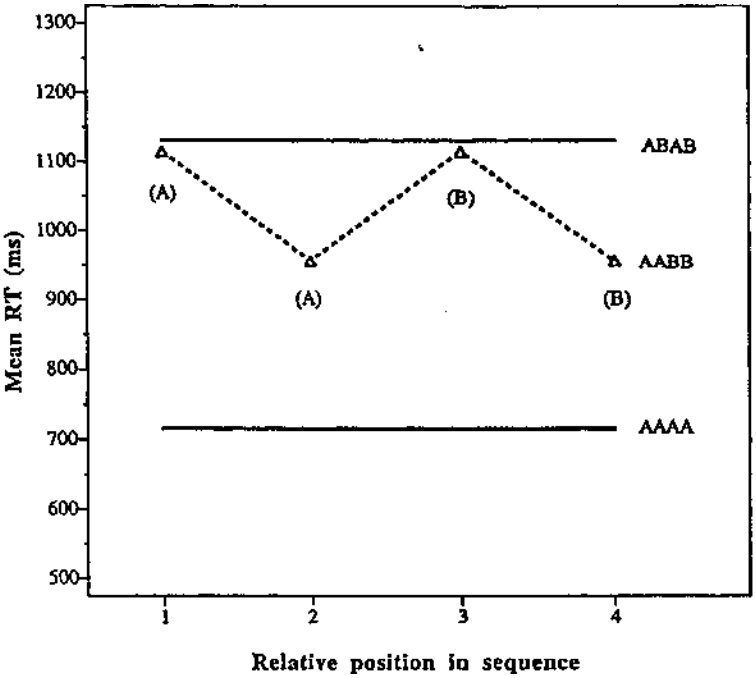

Figure 12.3 Reaction time (RT) results from an AABB task (Fagot 1994), redrawn in Pashler 1997. Subjects are faster on second performance of a given task, but still not so fast as in a pure block of trials.

As Rogers and Monsell (1995) point out, the alternation cost (difference between pure and alternating-task blocks) is likely to include several factors in addition to reconfiguration time. For example, there might be slowing due to the processing "overhead" needed to maintain and implement the intention to alternate. Because concurrent memory loads generally slow performance in reaction time tasks (Logan 1978), it seems reasonable to expect that holding onto a plan for alternating would impose a memory load of its own. In addition, differences in effort or arousal cannot be ruled out.

To help tease apart these factors, Rogers and Monsell used an "alternating-runs" procedure, wherein subjects performed first one task a number of times, then the other, and so forth. A pair of characters was presented on each trial, one a letter and the other a digit. Subjects either classified the letter as a vowel versus consonant, or the digit as odd versus even. Sometimes each task was performed twice in succession (AABB). The first response within a run of a given task ($\underline{A}A\underline{B}B$) was substantially slower than the second ($A\underline{A}B\underline{B}$), even at the long RSI. This was later confirmed by Fagot (1994) using the color and letter tasks described above. In AABB lists, subjects were required to perform the color task twice, the letter task twice, and so forth. Fagot also included pure blocks of trials and alternating (ABAB) blocks for comparison. As seen in figure

12.3, the first performance of a given task (AABB) was close to the ABAB blocks, but slower than the second performance (AABB), as in Rogers and Monsell's data. The second performance was still quite a bit slower than the pure block (AAAA), however, suggesting that the overhead cost is nontrivial.

In a further example of the stubbornness of the residual task switch cost at long RSIs, Goschke (chap. 14, this volume) allowed subjects 1.5 sec between two colored letters, each of which was to be classified by color or shape, and found responses were substantially slower when a different task had to be performed on the second letter.

Thus it appears, as Rogers and Monsell argued, that several factors play a role in the basic Jersild alternation effect. From the standpoint of conventional thinking in information-processing psychology, probably the most surprising of these factors is the switch cost, which persists even after ample time has been provided for reconfiguration. Some clues about the nature of this residual switch cost come from an additional experiment by Rogers and Monsell (1995), in which subjects performed a task four times in succession, then switched and performed the other task four times, and so forth. Performing a given task initially produced a substantial speedup for the second response, but over the next two responses, no additional improvement was detected (see figure 12.4). The authors concluded that the gain observed from performing the task once could not be attributed to "micropractice"—a small dose of the same optimization process that, over many trials, yields the familiar practice effect. After all, they reasoned, such a process could hardly reach an abrupt and final asymptote after one trial, as these data seem to show. The empirical basis for this conclusion has recently been challenged, however, by Salthouse et al. (1998), who had subjects switch tasks and then perform fairly long runs of a different task. They found RTs for the second trial within a run had still not reached a baseline in their data, and argued that Rogers and Monsell may have had insufficient experimental power to detect this continuing decline.

Cuing the Task Set

Control over task set is also illuminated by experimental designs in which the subject cannot tell which task to perform until a task cue is provided. Following in the footsteps of Shaffer (1965), Sudevan and Taylor (1987) had subjects perform one of two different tasks involving a digit. One task required classifying the digit as odd or even, while the other required classifying it as less than six or greater than five (bivalent mapping). The cue preceded the digit by an interval ranging between 400 msec and 4 sec. Responses became faster and more accurate as the interval was lengthened to about 2 or 3 sec. In his color/identity design, Fagot (1994) examined cue-target intervals ranging from zero to 4 seconds, and

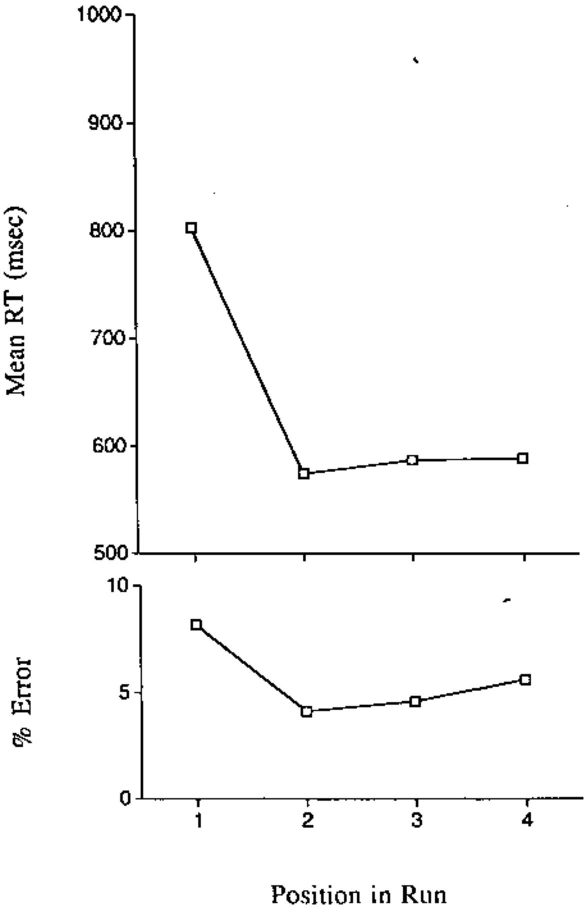

Figure 12.4 Reaction times (RT) and error rates in Rogers and Monsell 1995, experiment 6, as a function of position in a run of four trials (subjects performed one task four times, then switched to the other task; redrawn from Rogers and Monsell 1995, fig. 5).

found that the bulk of the benefit (over 200 msec) occurred over the range from 0 to 500 msec, with some further improvement out to about 1 sec; thereafter, performance was little changed. Other studies (e.g., Logan and Zbrodoff 1982) have also found a similar time course using cues that are helpful but not strictly necessary in performing the task. As Rogers and Monsell (1995) point out, one cannot directly derive an estimate of the time needed for reconfiguration based on these kinds of experiments, because at the shortest cue-target interval, the time needed to read and interpret the cue is presumably slowing responses, along with the requirement to reconfigure.

Recall that in the alternating-task procedure described earlier, the first response within a run of two successive instances of the same task is slower than the second response, even with an ample RSI. Based on that result, we would naturally expect that in the cuing procedure, no matter how long the cue-target interval, responses would be slower when the previous trial involved the other task. This is indeed the case. For example, Fagot (1994) presented task cues in blocks with a random or a fixed task sequence (either alternating or nonalternating). Even when subjects had four seconds to use the task cue, there was still a benefit of having performed the same task on the preceding trial; as expected, there was an additional benefit of having a fixed sequence.

In a clever recent study, Meiran (1996) cued subjects to respond to the vertical or horizontal position of a disk; the task varied from trial to trial within a block. Task cues (arrows pointing either up and down or left and right) appeared about 200 or 1,400 msec prior to onset of the imperative stimulus. Subjects were slower when they had to perform a different task from one trial to the next. This difference was substantially greater at the short cue-stimulus interval than at the long interval, but did not disappear at the longer interval. Meiran argued that the reduced task alternation effect produced by increasing the cue-target interval did not occur merely because lengthening this interval made the previous task more distant in time, reducing its impact by passive decay. When the interval between the previous response and the cue was decreased to make up for the increase in the cue-target interval, thereby holding the RSI constant, the longer cue-target interval still reduced the effect of a task switch. This strongly suggests that some, albeit incomplete, advance reconfiguration is indeed taking place.

Incompleteness of Reconfiguration

We have seen that in both the alternating-task procedure and the task-cuing procedure, providing subjects plenty of time to prepare reduces the cost of having to perform a task different from the one they just performed (in the bivalent situation), but it does not allow them to respond as quickly as if no switch of task had been required. This residual difference cannot be attributed to overhead cost because it appears also with the alternating-runs procedure (e.g., Rogers and Monsell 1995) as well as with the task-cuing procedure (e.g., Meiran 1996).

Why should there be a residual switch cost? Why is reconfiguration incomplete? De Jong (chap. 15, this volume) asked whether the residual switch cost stems from a constant slowing that appears on all trials or from a slowing that arises on only a fraction of the trials. Using the alternating-runs procedure of Rogers and Monsell, he had subjects classify colored letters according to either color or identity (consonant versus vowel). He found little evidence for a constant slowing component at the long RSI, and argued that incompleteness of reconfiguration is at least avoidable under certain conditions. His results may not rule out the possibility, however, that residual cost is always present, but imposes a delay whose magnitude varies from trial to trial.

Meiran (chap. 16, this volume) proposes that residual shift depends on a feature of certain switching designs not discussed thus far, namely, "ambiguity of responses," the use of an overlapping set of responses in the two tasks. As in his earlier experiments described above, Meiran used a design in which subjects respond to either the vertical or the horizontal position of a disk placed in one of four quadrants of the display. When the two tasks used the same two response keys (ambiguous responses), there was a positive residual task switch cost; when the responses were separate, the residual cost was reduced or absent.

According to Meiran, task preparation may involve not only the selective amplification or enabling of particular stimulus-response links, as depicted in figure 12.2, but also the selection of a response set, which can only be achieved by actually performing the task. This proposal is intriguing, and receives support from the data reported in this volume, although there are cases in the literature where residual task-switching costs have been observed even when two tasks did not involve "ambiguous" responses. For example, Fagot (1994) observed residual switch costs for mappings both with the same keys and with corresponding keys of different hands.

An alternative view of the residual cost of a task switch is that it results, not from the need to perform a time-consuming control process on the switch trial (as the authors described above have assumed), but from a prolongation on switch trials of the response selection process that happens on all trials. This prolongation, is caused by competition due to positive or negative priming of task sets or of S-R associations from previous trials on which the other task was performed. Such a view was first proposed by Allport, Styles, and Hsieh (1994), and a new version of it is presented by Allport and Wylie's chapter (chap. 2, this volume), to which the reader is referred for arguments and evidence. It seems clear from Allport and Wylie's work that there are carryover effects from recently performing the alternative task in response to the same stimulus or class of stimulus. What is not clear, however, is whether these carryover effects are sufficient to account for the dramatic drop in RT from the first to the second trial after a task swtich. Further, the notions of priming effects and control processes are by no means mutually exclusive.

Task Congruity Effects

The incompleteness of reconfiguration is revealed, not only by residual switch costs that persist despite long RSIs, but also by persisting effects of the purportedly disengaged mapping. Recall that Rogers and Monsell (1995) had subjects respond to either the letter or the digit in a letter-digit pair, using an alternating-runs procedure. The authors examined reaction times as a function of whether the irrelevant item in the pair would, according to the irrelevant (supposedly inactive) task mapping, yield the same response as that required on the current trial. The trial was called "congruent" when it did, and "incongruent" when it did not. There was a modest but significant tendency for slower responses on incongruent trials than on congruent trials, although responses trials with neutral stimuli were faster still. Similarly, in Fagot's color/letter design, where subjects responded either to color or to identity, responses were about 90 msec slower when the other feature was associated with a response inappropriate on the current trial. These congruency effects imply that the "competing task set is not entirely disabled" (Rogers and Monsell 1995, 216).

There is some controversy about whether the competing task set can be disabled when a sufficiently long RSI is provided. In their experiment 3, Rogers and Monsell found no significant reduction in the congruence effect (measured, as usual, in RTs) when they increased the R-S interval, although there was a marginally significant interaction in the error rates. Similarly, in three experiments, Fagot (1994) found only a weak reduction in congruency effects. By contrast, Meiran (1996), using his location button tasks, found a strong interaction, with congruence effects reduced but not eliminated. Finally, Sudevan and Taylor (1987) reported that congruence effects with their digit task disappeared at long cue-target intervals, while Goschke (chap. 14, this volume) reports having nearly eliminated the effect of task congruence with a long, unfilled RSI and after practice. Unhappily, then, the results run the full gamut from complete persistence of the congruence effect at a long RSI all the way to virtual disappearance. This issue remains to be sorted out.

Conclusions

Evidently, when subjects anticipate the need to perform a task incompatible with the one they just performed (as in the case of a bivalent list), whether this anticipation is based on the requirement to alternate (as in the Jersild paradigm and its spin-offs), or on the perception of a cue telling them to perform a task different from the one they just performed, some advance reconfiguration can occur, as depicted in figure 12.2. With the sorts of simple but arbitrary tasks studied in this literature, this reconfiguration usually seems to take under 0.5 sec when subjects have no other intervening task to perform. Reconfiguration may be accompanied by verbalization, usually covert, of the instructions for the upcoming trial.

The notion of advance reconfiguration illustrated in figure 12.2 seems to have some validity, but it misses important aspects of task switching. First, advance reconfiguration usually fails to eliminate the costs of having just performed a different task. Even with ample RSIs or cue-target intervals, subjects are still typically slower when they must perform a task different from the one they just performed (although Meiran's design reveals at least some exceptions). Actually performing a task once allows a significant amount of additional reconfiguration or tuning to take place. Rogers and Monsell refer to the tuning before first performing a task as the "endogenous component" of task preparation, and to that after performing the task as the "exogenous component." Although their data had suggested that exogenous reconfiguration is complete after one trial, subsequent data (Salthouse et al. 1998) suggest it may not be entirely complete until two trials.

Perhaps the most intriguing aspect of task switching is the lingering effect of the irrelevant mapping—the "task congruity effect." Not only advance reconfiguration, but indeed *all* reconfiguration accomplished up to the point of selecting a response in the new task seems incapable of wholly disabling the old mapping. While task congruity effects have on some occasions been observed to disappear with adequate preparation time, as noted above, more commonly they seem to persist, at least to some extent (an issue discussed in detail by Allport and Wylie, chap. 2, this volume).

12.2 DUAL-TASK PERFORMANCE

We turn now to the limitations that arise when people attempt to perform two different tasks at the same time. Our focus here will be on discrete tasks; with more continuous tasks, interference and switching are easily disguised for reasons that will emerge clearly below. Not surprisingly, limitations on simultaneous mental operations evidently arise at various different functional loci. Perceptual analysis of multiple stimuli often takes place in parallel, with capacity limitations sometimes becoming evident when perceptual demands exceed a certain threshold (Pashler 1997) although nonperceptual factors (such as statistical noise in search designs) often masquerade as capacity limitations (Palmer, 1995). These limitations appear largely, but probably not entirely, modality specific (Treisman and Davies 1973; Duncan, Mertens, and Ward 1997). Similarly, response conflicts arise when responses must be produced close together in time. These perceptual limitations are often most acute when similar or linked effectors are used, such as the two hands (Heuer 1985).

The most intriguing, and for the present topic most relevant, limitations arise in central stages of decision, memory retrieval, and response selection. Intuitively, most laymen assume that the cognitive aspects of two tasks can be performed simultaneously unless one or both are intellectually demanding. That this is not the case, however, is most clearly seen when people try to carry out two speeded but relatively simple tasks, each requiring a response to a separate individual stimulus. As Telford (1931) first observed, people almost invariably respond more slowly to the second stimulus when the interval between the two stimuli is reduced. Telford called this the "psychological refractory period" (PRP)

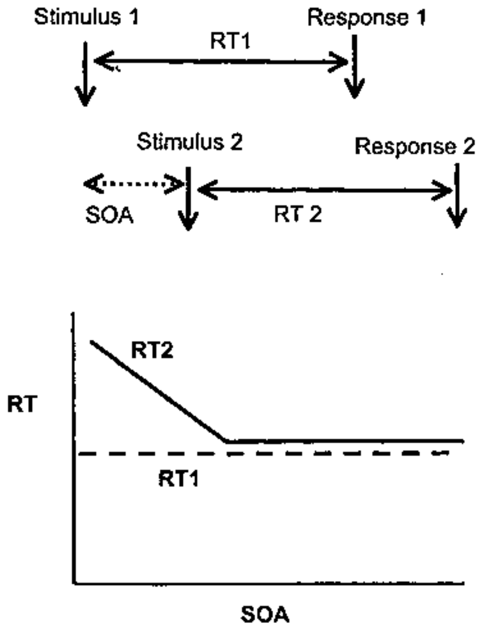

Figure 12.5 Schematic diagram of the psychological refractory period (PRP) design and idealized pattern of data (hypothetical numbers).

effect, by analogy to the refractory period of neurons. Though the analogy is probably not very apt, the label has stuck. In the PRP design, two stimuli (S1 and S2) are presented, their onsets separated by some stimulus onset asynchrony (SOA). The person makes a separate response to each stimulus (R1 and R2, respectively). Figure 12.5 (idealized data) shows the type of result usually obtained; the reaction time between S2 and R2 (RT2) grows as the SOA is shortened. Meanwhile, the reaction time between S1 and R1 (RT1) is usually relatively constant, although this depends on the instructions (see below). In some cases, the slope relating RT2 to SOA is as extreme as -1, which means that any reduction in SOA beyond a certain point merely increases RT2 by the same amount. To put it differently, presenting S1 and S2 closer together in time (once the interval reaches some minimum value) often fails to result in R2 being produced any earlier. Another important observation is that while processing required by the two tasks resists being "compressed" beyond a certain point, at short SOAs, the total time required to carry out both tasks (the interval between S1 and R2) is often substantially less than the sum of the times required to complete each task separately. In short, there is a saving in the total time for completing the two tasks, suggesting overlap in some aspects of processing.

The PRP effect has been observed in many different tasks, including simple reaction time (as in Telford's studies) and choice reaction time tasks (starting with Creamer 1963). Although early PRP experiments mostly used pairs of manual responses, sometimes made with the same finger, the effect can also be found when the two tasks involve completely different effectors. For example, PRP effects have been found with tasks combining manual and oculomotor responses (Pashler, Carrier, and Hoffman 1993), manual and vocal responses (Pashler 1990), manual and foot responses (Osman and Moore 1993), and vocal and foot responses (Pashler and Christian 1994). Thus it is not necessary for two tasks to use a common motor control system in order for a PRP effect to be observed. The PRP effect is also found when the two stimuli involve different sensory modalities. For example, Borger (1963) and Creamer (1963) found PRP effects with visual and auditory stimuli, as have many more recent researchers. It is not clear whether the PRP effect is greater when S1 and S2 are presented in the same modality; this is hard to determine because changes in input modality are typically confounded with differences in the compatibility of the task mapping.

Limits of the Psychological Refractory Period Effect

The PRP effect is very robust, but over the past twenty-five years or so, a number of exceptions have emerged. Greenwald and Shulman (1973; Greenwald 1972) found that the effect of SOA on second-task RTs was virtually eliminated when one task involved repeating a spoken word (shadowing) and the other involved a highly compatible visuomanual task. They hypothesized that "ideomotor compatibility," the fact that the stimulus mimics the feedback produced by the response, might be critical. Although McLeod and Posner (1984) demonstrated noninterference with combinations of shadowing and other tasks in ways that seemed consistent with this proposal, other research suggests ideomotor compatibility is probably not sufficient to eliminate interference. For example, Brebner (1977) devised a novel ideomotor-compatible task, requiring subjects to press a button in response to upward pressure from a solenoid located under the corresponding finger. When task 1 involved left-hand stimulation and task 2 involved right-hand stimulation, a clear-cut PRP effect was observed. Tasks requiring a saccadic eye movement toward a single spot, or even the generation of an eye movement in response to a single stimulus based on its color, seem not to generate PRP effects (Pashler, Carrier, and Hoffman 1993). Visuomanual tasks with very high spatial stimulus-response compatibility may also be free of central interference (Koch 1994). At present, then, the conditions under which the PRP effect disappears are not well characterized. Indeed, it seems that dual-task interference in pairs of punctate tasks can be eliminated only with tasks that are, intuitively speaking, extremely natural and easy. Whether the critical factor is the existence of prewired neural circuits that bypass central machinery, a high degree of practice, or some combination of these factors remains unknown. Perhaps the more significant point is

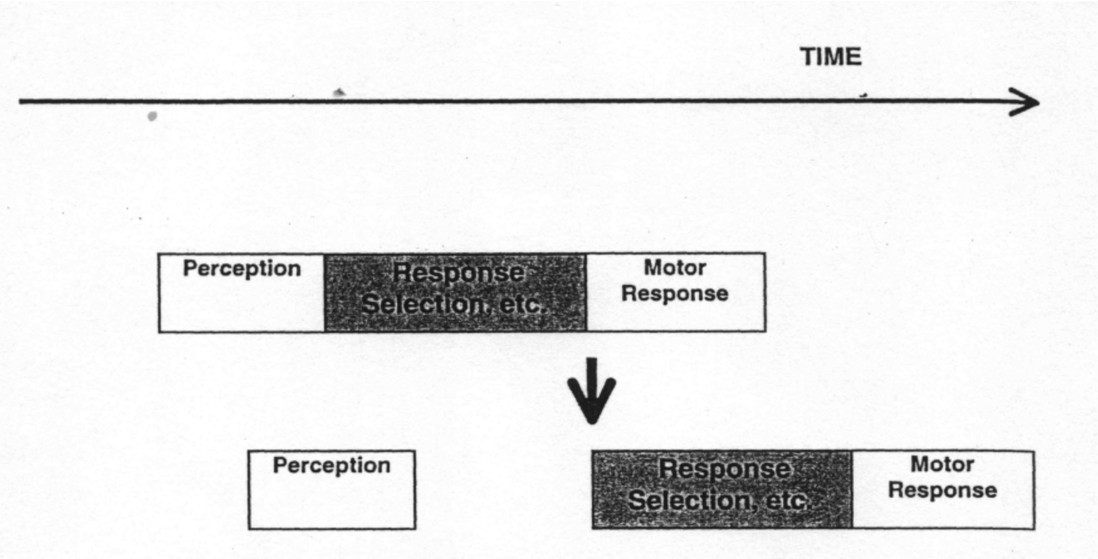

Figure 12.6 Central bottleneck account of the psychological refractory period (PRP) effect.

that it is easy to find tasks with minimal cognitive demands that produce robust PRP effects.

Sources of Dual-Task Slowing

Based largely on observations of PRP interference even where there is no overlap in stimulus or response modality, Welford (1952, 1980) proposed that dual-task slowing arises from a bottleneck in what he called "stimulus-response translation"—in more modern parlance, the stage of "response selection." The basic idea is illustrated in figure 12.6. According to this hypothesis, each task is composed of three broad stages (perception, response selection, and response execution); any stage of task 1 can overlap any stage of task 2, except for the shaded stage of response selection: while one response is being selected, selection of the other response must wait. As formulated, however, the hypothesis does not say what should happen in tasks more complicated than choice RT, where one often encounters mental operations that do not obviously fall into any one of the three categories.

From this account, one can derive very specific predictions for the results of dual-task experiments in which different stages of task 1 or task 2 are selectively prolonged. Increases in the duration of stages of task 1 up to and including the shaded stage should, at short SOAs, propagate and slow task 2 as well as task 1. Increasing the duration of the post-bottleneck stages of task 1, on the other hand, should slow only task 1, regardless of the SOA. Increasing the duration of stages in task 2 prior to the bottleneck should correspondingly slow the second response at long SOAs. At short SOAs, on the other hand, there is "slack" because the response selection in task 2 is not waiting for completion of perception in task 2, but rather for the completion of response selection in task 1. The result of the slack is that, at short (but not at long) SOAs, the perceptual slowing should cease to affect RT2. The prediction, then, is that manipulations of the prebottleneck processing stages in task 2 should exhibit underadditive interaction with SOA (see Jolicoeur, Dell'Acqua, and Crebolder, chap. 13, this volume, for further details and examples). Lengthening the duration of stages at or after the shaded portion of task 2, on the other hand, should always slow R2 to the same extent, regardless of SOA.

These predictions have been confirmed in many experiments involving fairly elementary choice RT tasks (for a review, see Pashler 1997). The predictions are distinctive in the sense that they not only favor the central bottleneck, but also rule out accounts that would place the bottleneck ear-lier or later in the sequence of processing stages. Several of the results also seem unfavorable to graded capacity-sharing models, especially the fact that increases in first-task response selection difficulty have at least as large an effect on RT2 as it has on RT1 (e.g., Broadbent and Gregory 1967). If task 1 were being performed with depleted capacity, and the manipulations increased the capacity required to carry out the stage in question, one would expect to see a greater effect on RT1 than on RT2 (see Pashler and Johnston 1998 for discussion).

Much recent work within the bottleneck framework has focused on the question of exactly which processes are subject to this limitation, and which are not. Manipulations of the duration of sensory processing in task 2 (e.g., contrast) show the underadditive pattern indicating that the stages affected are not subject to the bottleneck (Pashler 1984; De Jong 1993). Johnston and McCann (forthcoming) degraded letters by making them very squat or very narrow without altering stroke widths and contrast. In another experiment, they altered the tilt of strokes composing the letters (for instance, rotating the diagonal segments in the letter *A* inward so that the character looked something like a teepee). At long SOAs, these distortions slowed RT2 by about 30 msec. At short SOAs, however, RTs for distorted and undistorted were indistinguishable, suggesting absorption into slack. It seems likely, therefore, that letter identification, not merely visual feature extraction, can occur on task 2 while critical stages of task 1 are under way. On the other hand, when perceptual processing demands on task 2 include not just identifying stimuli, but performing additional manipulations such as mental rotation or comparisons, these operations are usually subject to the central bottleneck (Ruthruff, Miller, and Lachman 1995).

Recent evidence suggests that, not merely the planning of actions based on task-mapping instructions or difficult perceptual manipulations, but memory retrieval overall is subject to queuing. Carrier and Pashler (1996) combined a manual response to a tone (task 1) with paired associate retrieval cued by a visually presented word (task 2) in a PRP design. The duration of the memory retrieval was manipulated by varying the amount of practice subjects had carrying out any particular retrieval. Second-task RTs were, not surprisingly, faster for better-learned pairs. In the dual-task situation, this difference appeared additive with SOA (Carrier and Pashler 1996). Following the logic described above, this implies that memory retrieval was postponed by first-task processing and refutes the claim that only the execution of the motor response is delayed. The latter point seems especially clear because of the greater difficulty of task 2 compared to task 1. In the short (50 msec) SOA condition, subjects responded to the tone about 600 msec after it was presented; the paired-associate task was far more challenging, however: on average, the paired-associate response did not occur until about 1,100 msec later. If all interference were response related, it is hard to see what could be postponing a second response so temporally remote from the first. The results are to be expected, however, if one assumes that the central bottleneck encompasses both response selection in task 1 and memory retrieval in task 2 (and perhaps response selection as well, if that is a separate stage in this sort of task).

It seems to me a reasonable conjecture that the inability to select two responses at the same time, which is apparent in choice RT tasks (Welford's response selection bottleneck), may be just a special case of a broader constraint, namely, that two retrievals cannot be carried out at the same time. Within the confines of the choice RT experiment, it is an action plan that is to be retrieved, whereas in other situations, it may, for example, be a word or concept or episode. While the proposed constraint can be expressed very simply, it stands in great need of explication. For example, what is meant by "two retrievals"? If two stimuli are presented, each associated with the same single response, does the lookup of that single response based on the two stimuli constitute two retrievals or one? In choice RT tasks, two redundant stimuli produce what Miller (1982) calls "coactivation," a particularly strong form of parallel processing. The same is almost surely true of more time-consuming memory retrieval operations. What about one stimulus associated with two responses? Timothy Rickard and I (Rickard and Pashler 1998) trained subjects in one phase of training to associate each item on a list of ten words with a corresponding verbal paired associate, and then, in a second phase of training, to associate each item on the same list with a manual response.[4] In a final test phase, subjects were sometimes instructed to carry out both retrievals at once. Whichever response was produced second had on average a latency that was about twice as long as the single-task control. Other aspects of the data also argued that the retrievals were carried out sequentially. Thus, for the purpose of the proposed constraint, it is the number of outputs, not the number of inputs, that determines whether a single retrieval or multiple retrievals are required.

The term *retrieval* also needs clarification. A priori, one might have described letter identification, for example, as involving the retrieval of the letter identity corresponding to a visually presented character. Yet I have argued that object identification is not subject to the bottleneck. What differentiates retrieval from classification or identification? At this point, the answer must be vague: it seems that the operations subject to queuing involve retrieving some mental contents that are distinct from the input in that they are not an internal description of the input but some separate contents. Sharpening up this description will require at the very least testing a broader range of different types of retrievals in different dual-task contexts; conceivably, it will also require a better understanding of the neural substrates of these processes.

Strategic Interpretations

The apparent inability to execute the central stages of even fairly easy tasks concurrently is surprising from both an intuitive and a computational standpoint. It has recently been argued that postponement of central processing in the PRP design stems not from a fundamental inability to carry out the two tasks at the same time, but rather represents a strategic response to the explicit or implicit demands of the experiment. This idea has been developed in detail by Meyer and Kieras (1997), who proposed an ambitious theory of human performance ("executive process interactive control" or EPIC), discussed in detail by Kieras et al. (chap. 30, this volume). According to EPIC, there are no intrinsic limitations whatever in the ability to select responses or carry out memory retrievals concurrently. There are, however, structural limitations in the initiation and execution of responses. In addition, postponement of central processing (i.e., queuing of processing stages) may occur whenever subjects perceive this to be advantageous.

Why would subjects adopt a queuing strategy in a dual-task design when doing so means responding more slowly in one or both tasks? As Meyer and Kieras note, in many PRP experiments, subjects have been told to produce Rl as fast as possible (and even, in a few cases, to produce Rl before R2). Primarily, this has been done in order to avert the "grouping" strategy that people naturally fall into, whereby Rl is buffered and then emitted shortly before R2 (Borger 1963; Pashler and Johnston 1989, exp. 2). Given a strong emphasis on first-task speed, subjects might choose not to select the two responses in parallel because doing so might result in responding to task 2 before task 1.

One obvious question, then, is what happens when there is no emphasis on the speed of the first task and subjects try to respond to both tasks as quickly as possible. A number of studies that did not emphasize first-task speed have nevertheless shown evidence of central postponement. For example, in Carrier and Pashler 1996, even though subjects were not told to emphasize the speed of the first response, both slowing of R2 and postponement of central processing were observed. Similarly, in one of their experiments, Ruthruff, Miller, and Lachmann (1995) did not emphasize first-task speed but nonetheless found evidence of a central bottleneck.

There are also some other, rarely cited studies in which investigators looked at performance of two serial choice RT tasks, where subjects are instructed to respond to a train of signals in each task, rather than to two discrete signals, as in the PRP design. Here the order of responses is entirely up to the subjects, who simply attempt to achieve as much "throughput" as possible in each task. Gladstones, Regan, and Lee (1989), for example, had subjects perform serial tasks paced by the experimenter (e.g., pressing a key in response to the position of a light and pronouncing a letter in response to the color of a light). In some conditions, subjects performed just one such task, whereas in others, they performed two concurrently. The total rate at which information was processed summed over the two tasks (which corresponds roughly to the total number of responses in either task per unit time) was the same whether one task was performed or two. This was true even after considerable practice, and regardless of whether the tasks used the same or different input and output modalities. Similar findings were reported by Fisher (1975a,b) and Schouten, Kalsbeek, and Leopold (1960). Although, following Meyer and Kieras (1997), some interference might be expected due to conflicts in the initiation of responses, a bottleneck confined to response-related processing should allow a dramatic increase in total throughput rate to be achieved when two tasks are performed, instead of one.

My colleagues and I recently carried out other kinds of studies using discrete tasks to examine whether central queuing is strategic in origin. In one study, Eric Ruthruff, Alwin Klaassen, and I instructed subjects to perform two tasks and group the responses close together in time, a requirement subjects find quite natural. One task required judging whether a figure was a normal or a mirror image letter and making a corresponding keypress response. The other task, which could be performed more quickly, involved discriminating between a single 17 msec tone and a rapid-fire sequence of two 17 msec tones separated by 50 msec, with a vocal response (saying "one" or "two"). The first tone and the letter began simultaneously.

The instruction to group the two responses obviously does not provide any incentive to perform one task before the other. If there is no interference between the decision or response selection phases of the two tasks, the response should almost always be selected more quickly in the easier task, normally the tone judgment. Thus the grouped response should only be a bit slower than the response for the letter task alone, due to occasional trials in which the letter task happens to take longer than the tone task, plus any cost associated with grouping. In fact, there was very substantial slowing of mean RTs (1,475 msec for the dual-task grouped response, compared to 917 msec for the letter task alone). Monte Carlo simulations disclosed that this slowing could not be accounted for by the fact that the tone task was occasionally slower than the letter task. It is also not likely to reflect extra time taken to produce a grouped response; costs of producing grouped responses can be assessed directly, and prove negligible (e.g., Pashler and Johnston 1989, exp. 2).

As a further test, the difficulty of response selection in the easier task was varied: in compatible blocks, subjects responded by saying "one" to the single tone pulse, and "two" to the two pulses; in incompatible blocks, the mapping was reversed, producing about 200 msec of slowing. If central processing on the easier task were carried out in parallel with central processing on the harder task, much of the slowing of the tone task should be absorbed in "slack," and thus have minimal effect on the time to produce the grouped responses. In reality, compatibility had at least as large an effect on the grouped response in the dual-task context as it had on performance of the tone task by itself. Thus the whole pattern of results in this experiment favors the idea that central queuing was occurring in a situation where parallel processing would clearly have been advantageous.

In another recent study, Levy and I required subjects to make a three-alternative button-push response to the color of a large disk presented on a monitor screen, and to make a vocal response to its position (saying "one", "two" or "three" for left, middle, or central position). Here, rather than using grouping, we provided explicit payoffs designed to promote parallel processing and to place equal emphasis on the speed of each task. On blocks where both stimuli were presented, average reaction times for both tasks exhibited substantial slowing. Again it appears that encouragement to prioritize one task more than the other is by no means a necessary condition for dual-task interference to occur.

12.3 RELATING DUAL-TASK INTERFERENCE AND TASK SET

Having very briefly and selectively reviewed some of the main phenomena in the area of task switching and central limitations in simple dual-task performance, let us consider possible relations between the two topics. The research on dual-task interference bears on the issue of task set and task switching in several interesting respects. Two of these will be discussed here. The first is a very broad question of cognitive architecture: Do the phenomena of task set reconfiguration and dual-task interference (and specifically the sort of central queuing argued for in section 12.2) singly or jointly imply the existence of a "central executive" or "supervisory attention system"? The second question is narrower: Does the bottleneck itself reflect a limitation in task set, and perhaps the same limitation as is responsible for task-shifting costs, in which case the phenomena of dual-task queuing and task switching might really be one and the same?

Many writers have assumed that cognitive control requires the existence of a specific controlling mechanism whose function is to program (other) cognitive machinery. As discussed in several chapters in this volume, this controlling function is often associated with the frontal lobes or specific parts thereof. Several well-known theoretical frameworks in cognitive psychology, such as Baddeley's dissection (1986) of working memory and Norman and Shallice's theory (1986) of attention and control, famously invoked the idea of a "central executive." For present purposes, we can put aside the common criticism that invoking an executive as an account of mental control creates a sort of infinite regress (does the executive contain its own executive?). Rather, let us simply ask whether the phenomena of set and dual-task interference provide any sort of evidence for such a conception.

As several authors have pointed out (e.g., Allport 1987; Monsell 1996), the alternative is a scheme in which executive control emerges from the interaction of the very same machinery that ordinarily carries out the mental processes being controlled. The brute phenomena of executive control (e.g., that we can decide to perform one task or another; that verbal instructions can, if their recipient chooses to comply, completely determine which stimuli evoke which responses) emphatically do not require the existence of machinery dedicated for the purpose of control. Mutual competition between distributed mechanisms for the control of thought and action may well account for task set-switching phenomena. Indeed, work on "multiagent planning" in artificial intelligence suggests such a mechanism is capable of much more than that (e.g., Suarez, Winstanley, and Griffiths 1998). Furthermore, some of the phenomena of task set described above, such as the need to perform at least one trial of a new task in order to fully reconfigure processing machinery for that task, seem slightly more congenial to a distributed control architecture than to the notion of a distinct executive mechanism.

It is also commonly suggested that the idea of an all-or-none processing bottleneck (particularly a single bottleneck that spans diverse cognitive contents, as argued for above) naturally implies or at least suggests the existence of a single mechanism that carries out whatever cognitive operations are subject to queuing. Noting this, some writers (e.g., Kinsbourne 1981) have pointed out that the notion of a single-channel bottleneck seems hard to reconcile with the highly distributed processing that characterizes the human cerebral cortex.

It is certainly true that one very natural explanation for obligatory queuing of any given operation is the possibility that there is only a single device capable of carrying out the operation. That may not be the only explanation, however, let alone the correct one. Consider, for example, recent studies of processing bottlenecks in commisurotomy ("split-brain") patients. If the central bottleneck described above has a defined cortical locus, split-brain patients should show no PRP effect whenever each task is confined to a separate hemisphere (assuming they are capable of performing the tasks under such conditions). However, using lateralized stimuli and responses, Pashler et al. (1994) observed relatively normal performance and a relatively normal PRP effect in four split-brain patients. We concluded that the queuing underlying the PRP bottleneck must have a subcortical source because connections at these brain levels remain intact in split-brain patients (but see Ivry and Haseltine, chap. 17, this volume, for another view based on later studies conducted with one of these patients). It seems very unlikely that a brain stem mechanism would be responsible for actually carrying out memory retrieval and response selection. The natural alternative, then, is that the operations subject to queuing are themselves distributed and subcortical mechanisms trigger or control the queuing.

Is Queuing a Consequence of Task Set Limitations?

Is it possible that difficulties in selecting two responses at the same time (resulting in the PRP effect) stem from an inability to simultaneously maintain the task set for the two separate tasks? Although this idea has been suggested from time to time (e.g., Gottsdanker 1980), such a reduction seems hard to reconcile with the task-switching phenomena described earlier in this chapter.[5] Recall that in the Jersild paradigm, people usually incur only a fairly modest cost (and sometimes none at all) in shifting from one task to another so long as the mapping is univalent (i.e., where no stimulus is ever mapped onto different responses in the two tasks). Because, in the typical PRP task, the stimulus sets for the two tasks are nonoverlapping, the problem of concurrent task set maintenance should be comparable to that found with the univalent Jersild task, not with the bivalent task. Based on the results described earlier, one would therefore expect to find only a fairly modest slowing, presumably because both tasks sets can simultaneously coexist. Because the PRP effect often reaches several hundred milliseconds, presumably this concurrent maintenance problem cannot be the whole source of it.

On the other hand, one need not rely on indirect inferences; the concurrent maintenance contribution to PRP slowing can be assessed fairly directly, with a control seldom used until recently, by introducing to the PRP experiment blocks in which subjects prepare for both tasks, but are presented only one stimulus and are unable to predict which one this will be. In one unpublished study, Eric Ruthruff and I had subjects make a verbal response to a color patch, a manual response to a tone, or both. In the "or" task, subjects performed one task or the other, but not both (only one stimulus was presented). The "and" task was basically a PRP task with a zero SOA. There was some slowing in the "or" task compared to pure task blocks, but much more slowing on top of that in the "and" task.

The preparatory limitation responsible for the slowing in the "or" task as compared to a pure single task is likely to be responsible for slowing found in various single-task designs, as Gottsdanker (1980) pointed out. In a choice RT task, a greater number of stimulus-response (S-R) pairs is associated with longer RTs (Hick 1952)—an effect that depends chiefly on the number of alternatives subjects must prepare for, rather than the number of different alternatives they were exposed to during the current block of trials (Dixon 1981). Presumably, the need to prepare more S-R "links" means that each link cannot be prepared as fully, causing performance to be slowed (Gottsdanker 1980; Logan 1978). It is not merely the number of links that matters, however; the more conceptually cohesive the set of stimuli mapped onto any single response, the faster the task can be performed (Greenwald, McGhee, and Schwartz 1998; Seymour 1973). What is not clear is how preparatory costs should be understood. For example, does poorer preparation for larger or more heterogeneous mappings reflect more time having elapsed since a given link was prepared, or is "preparatory capacity" subject to continuous sharing, as proposed by Gottsdanker (1980)?

A Modified Reduction Hypothesis

Even though dual-task slowing is not reducible to the preparatory limitation for the reasons just discussed, one could still try to explain the PRP effect in terms of a limitation in task set. Consider the following hypothesis. In the "or" task experiments just described, the response selection module might not be preset at all, or it might be set in a "neutral" fashion. The shift from this unprogrammed state to the appropriate task set might occur very quickly, producing only a minor cost. Suppose, counter to what we have been assuming throughout this chapter, that, in the PRP design, despite a univalent mapping, the first task set must be disengaged and the second task set loaded before the second task can be processed. To explain why the dual-task case ("and" task) produces more slowing than the unknown single-task case ("or" task), one merely has to suppose that the response selection machinery cannot be reprogrammed while it is being used. This does not seem like an unreasonable supposition. The only problem is that because this account presumes that task set reconfiguration is necessary even with univalent mappings, it fails to explain why bivalent lists exhibit so much more alternation cost than univalent lists, although, with some ingenuity, it could probably be made to explain this as well.

Fortunately, however, we do not need to rely on such arguments. What would provide a critical test of the hypothesis that the bottleneck reflects a limitation in maintaining the set for each task? If the bottleneck reflects an inability to prepare the two task mappings simultaneously, then it should disappear when two or more tasks use the same mapping. That is, if the stimulus-response mapping rule remains fixed, and several stimuli must be processed, parallel central processing should be possible, unlike in the normal PRP case. One possible test of this claim would use a PRP task in which two distinct stimuli are presented and the response rule is the same.[6] Another method in which the mapping remains constant but subjects attempt to perform more than one task at the same time is the serial RT task, where subjects respond to a whole string of stimuli. In a recent study, we had subjects carry out a self-paced serial task, with and without preview (Pashler 1994). Letters unfolded from left to right, and subjects made a button-push response to the identity of each letter (four possible keys and four possible letters); ten letters unfolded, so that at the completion of the trial, there were ten letters on the screen and subjects had made ten responses. In the no-preview condition, the experiment began with the presentation of a single letter; stimulus $n + 1$ was presented as soon as subjects responded to stimulus n. In the preview condition, the experiment began with two letters on the screen; stimulus $n + 1$ was presented on the response to stimulus $n - 1$. Due to the preview, subjects could potentially begin processing stimulus $n + 1$ while still processing stimulus n. Is this logical possibility also a psychological possibility?

The rate of responding in the preview condition was greater than in the no-preview condition. First noted by Cattell (1886) and confirmed by Leonard (1953), this finding strongly suggests that *some* overlap of processing stages does indeed occur in the preview condition (as it does in the conventional PRP situation, too; see figure 12.7). The key question was whether the response selection stages associated with successive stimuli could overlap. To answer this question, several different task difficulty manipulations were used: targeting perception, response selection, and response production. When the mapping was made less natural, thereby increasing response selection duration (the manipulation was applied for the whole list of ten stimuli), the time between each response in the run was increased. The slowing was the same with or without preview. On the other hand, when perceptual processing was made more difficult, the time between the first stimulus and the first response lengthened, but the rate of responding thereafter was virtually unaffected. The results can be summarized by saying that response selection (but not perception or response production) seems to be rate limiting for serial performance even when stimuli are presented well before they are needed. Evidently, only one response can be selected at a time even if the rule for selecting responses does not change.

If the need to select new responses without any need to change task set is sufficient to produce response selection queuing, it seems gratuitous (or at least unparsimonious) to attribute the bottleneck in selecting com-

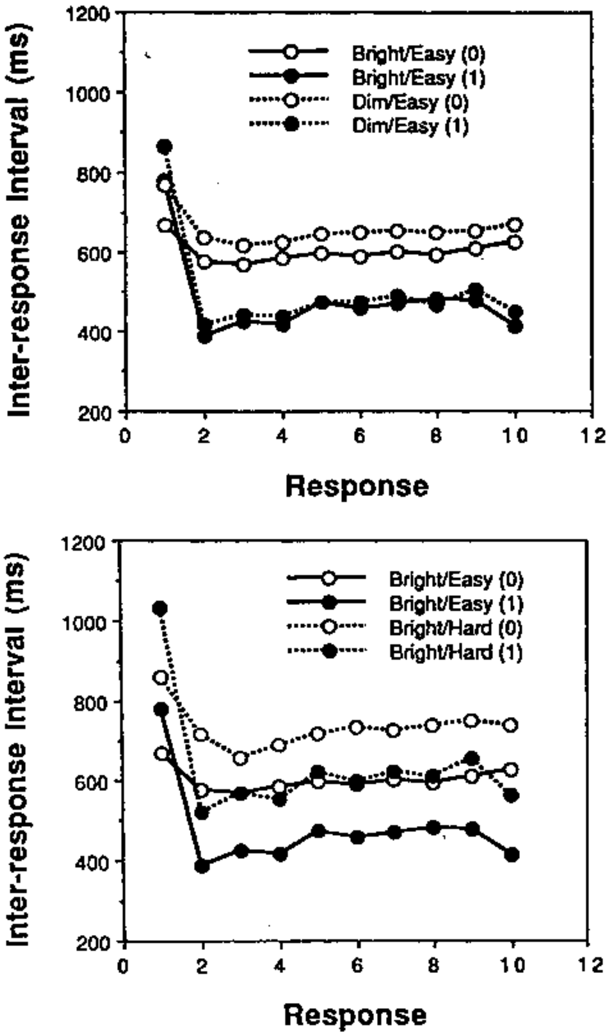

Figure 12.7 Effects of preview, stimulus quality and S-R compatibility on serial reaction time task. From Pashler 1994.

pletely distinct responses to an inability to maintain nonoverlapping (univalent) mappings simultaneously prepared. In view of this finding, plus the minimal cost of shifting in univalent lists (Jersild and others), it seems likely the limitation on carrying out two response selections at once cannot be reduced to a limitation on maintaining the two task sets at once. Presumably, because the mappings are univalent, the response selection module is loaded with both mappings (although not without cost, and not necessarily to the same degree at all times throughout the trial). That would suggest that the order of task performance in the PRP situation is probably not preplanned, a view that has been challenged by De Jong (1995). Logically speaking, there is no contradiction between saying that the two task mappings are simultaneously loaded and saying that the order of processing is planned or anticipated, although De Jong's evidence for preplanning of order involved tasks with two manual responses, and may therefore represent a rather special form of response selection.[7]

Alternative Explanations for Bottlenecks

We have considered two possible reasons for why a bottleneck might arise in the process of action planning (and, it was suggested above, memory retrieval as well). One explanation suggested that the bottleneck reflects strategic choices in scheduling mental operations, rather than a structural limitation: the other, that it reflects a limitation in simultaneously maintaining the two mappings in an active state. The evidence described above, although not fully conclusive, suggests that neither of these explanations is likely to be correct. If so, how else might one account for this puzzling limitation?

One intuitively very appealing idea, proposed by Allport (1987, 1993) and endorsed by De Jong (chap. 15, this volume), is that a bottleneck in planning might serve a positive function of preventing incompatible actions, thus maintaining the overall coherence of our behavior. The PRP effect, which appears as an obstacle to optimal performance within the contrived constraints of the dual-task experiment, might therefore be adaptive—in computer parlance, a "feature, not a bug." This proposal does not explain, however, why even time-consuming memory retrievals should be subject to queuing, as argued above. Nor, as formulated, does it specify exactly what sort of incoherence is meant to be prevented by queuing. One idea might be that preventing unrelated actions from being selected simultaneously would prevent the simultaneous execution of motor responses created by different action plans. This, it might be argued, would help maintain the coherence of behavior because a single planning operation will seldom (one might assume) generate behaviors that are mutually disruptive. The problem with this idea is that we are actually quite capable of simultaneously executing responses reflecting two or more independent planning operations. Casual observation of ordinary human activities reveals many examples. In a cafe, for example, a patron will lower a coffee cup while simultaneously beginning to speak; in a store, a clerk greets a customer while simultaneously putting the previous customer's groceries in a bag. It seems far-fetched to suppose that the speech and the hand movement, or the greeting and the hand movement, result from a single plan. These informal observations are confirmed by objective data. Van Galen and ten Hoopen (1976), for example, had people pronounce multisyllabic words in response to a letter and then make a button-push response to a second letter that followed soon after. The button-push response often occurred while the vocal response was still in progress; when this happened, there was no detectable interference.

One might suggest that what the brain is engineered to avoid is not the overlapping execution of independently selected responses, but rather the planning of an action that would terminate or disrupt a previously selected action. Such a constraint might, in de Jong's words (chap. 15, this volume), "protect task performance in progress from interference." Here again, there is little reason to believe that the constraint envisioned really exists. People can cancel actions that have just been launched, even when these are highly practiced. For example, Logan and Burkell (1983) showed that skilled typists could rapidly stop typing when an auditory stop signal was presented. In simple terms, action planning and the earliest stages of execution are not "ballistic." If they were, it might lend a certain form of coherence to our behavior, but probably a sort of coherence we should be glad not to possess.

The obvious alternative to accounts that view queuing as a positive benefit are accounts that claim the computational requirements of parallel retrieval would exceed available resources. This is somewhat puzzling, though, in view of the rather elementary kinds of task mappings that elicit queuing. The possibility of cross talk between tasks may help explain the ubiquity of queuing, if not quite as directly as some writers have supposed. Because similarity of tasks seems *not* to be a necessary condition for dual-task interference or queuing, attributing dual-task interference to content-specific cross talk within a given task combination seems rather unpromising (Pashler 1997). It is possible, however, that the system is wired up to require queuing as a general policy (conceivably one that can be overcome with sufficient practice) to prevent cross talk from unpredictably degrading performance in certain cases. Such an account seems consistent with several findings described earlier, including the proposed unity of limitations in action selection and memory retrieval, and the evidence from split-brain patients that anatomically distributed processing can be subject to queuing.

Open Questions

The study of task set is in its relative infancy, and the suggestions offered here about how we might relate task set to dual-task limitations are modest and preliminary. Many very basic questions remain to be addressed. One obvious question is whether the process of task reconfiguration itself can be carried out in parallel with another task. Goschke (chap. 14, this volume) finds that people are able to achieve the usual (partial) degree of reconfiguration if required concurrently to verbalize a description of the task they are about to perform. On the other hand, producing an irrelevant verbalization interfered with reconfiguration. What is not clear is whether carrying out an unrelated nonverbal task would interfere. This issue seems quite amenable to chronometric study.

Another open question is how the concepts useful in thinking about arbitrary choice reaction time tasks that have been the focus of the research described here might generalize to the more ordinary activities of everyday life. In activities like driving and conversing, one may speak of "task schemata" or "goals," but the notion of "mapping" seems inapt or contrived. Unfortunately, the implications of many of the concepts described here for such tasks remain to be clarified. This statement is not intended as a criticism of researchers who have, reasonably enough, started by studying relatively tractable cases. One area where some steps have been taken toward greater "ecological validity" is bilingual lexical production. Several investigators have given bilingual subjects cues telling them to name stimuli such as numbers in one language or another, and examined the effects of RSI and related variables. Thus far, the results with this task seem encouragingly similar to those found with non-linguistic laboratory tasks described above (MacNamara, Krauthammer, and Bolgar 1968; Meuter and Allport 1999). It is to be hoped that further efforts to examine tasks of this sort, as well as classic laboratory tasks, may shed greater light on the issues of task control and dual-task performance.

NOTES

This work was supported by National Institute of Mental Health grant 1-R01-MH45584 and by National Science Foundation grant SBR9729778.

1. Why preview should produce a switch benefit remains an open question. Conceivably, people can overlap more of the processing of each successive task when the mapping is changing.

2. In Vicentized distributions, the values for different percentiles are determined separately for each subject, then averaged across subjects; the results represent the typical shape of individuals' distributions, even if their speed of responding differs greatly.

3. In some cases (e.g., Rogers and Monsell 1995, exp. 4), a significant switch cost has been found with univalent lists that use compound stimuli, where the irrelevant stimulus was neutral (i.e., associated with no response).

4. The stimulus terms were color names and the verbal response terms were digits. During testing, single- and dual-task blocks were interspersed.

5. Note that the issue here is not whether the PRP effect arises merely as a consequence of temporal uncertainty about when S2 will arrive. This idea is clearly refuted by the finding that when the temporal parameters are unchanged, but subjects need not respond to S1, no PRP slowing occurs (e.g., Pashler and Johnston 1989).

6. One would naturally assume that sensory- or effector-specific interference would potentially contaminate such a study. If, however, the duration of central processing substantially exceeded that of more peripheral processing, reuse of the same sense and effector mechanisms should make very little difference; this deserves testing.

7. Manual response selection may ordinarily choose a spatial location, rather than a finger. If both a left-hand response and a right-hand response must be selected, the potential set of spatial locations may be unwieldy. As a strategy, the response-selection machinery might therefore choose within-hand spatial locations for each hand in turn, requiring a planned order (Pashler 1990). If this explains De Jong's findings, evidence for preplanning ought to disappear when one task is manual and the other vocal.

REFERENCES

Allport, A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In H. Heuer and A. F. Sanders (Eds.), *Perspectives on perception and action,* pp. 395-419. Hillsdale, NJ: Erlbaum.

Allport, A. (1993). Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years. In D. Meyer and S. Kornblum (Eds.), *Attention and Performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuro-science,* pp. 183-218. Cambridge, MA: MIT Press.

Allport, A., Styles, E. A., and Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umilta and M. Moscovitch (Eds.), *Attention and Performance XV: Conscious and nonconscious information processing,* pp. 421-452. Cambridge, MA: MIT Press.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Biederman, I. (1973). Mental set and mental arithmetic. *Memory and Cognition, 1,* 383-386.

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

Brebner, J. (1968). The search for exceptions to the psychological refractory period. In S. Dornic (Ed.), Attention and Performance VI, pp. 63-78. Hillsdale, NJ: Erlbaum.

Broadbent, D. E., and Gregory, M. (1967). Psychological refractory period and the length of time required to make a decision. *Proceedings of the Royal Society of London B168,* 181-193.

Carrier, M., and Pashler, H. (1996). The attention demands of memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,*1339-1348.

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

Creamer, L. R. (1963). Event uncertainty, psychological refractory period, and human data processing. *Journal of Experimental Psychology, 66,* 187-194.

Do Jong, R. (1993). Multiple bottlenecks in overlapping task performance, *journal of Experimental Psychology: Human Perception and Performance, 19,* 965-980.

De Jong, R. (1995). The role of preparation in overlapping-task performance. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48A,* 2-25.

Dixon, P. (1981). Algorithms and selective attention. *Memory and Cognition,* 9, 177-184.

Duncan, J., Martens, S., and Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature, 387,* 808-810.

Fagot, C. (1994). Chronometric investigations of task switching. Ph.D. diss., University of California, San Diego.

Fisher, S. (1975a). The microstructure of dual-task interaction: 1. The patterning of main-task responses within secondary-task intervals. *Perception, 4,* 267-290.

Fisher, S. (1975b). The microstructure of dual task interaction: 2. The effect of task instructions on attentional allocation and a model of attention-switching. *Perception, 4,* 459-474.

Gladstones, W. H., Regan, M. A., and Lee, R. B. (1989). Division of attention: The single-channel hypothesis revisited. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 41 A,* 1-17.

Gottsdanker, R. (1980). The ubiquitous role of preparation. In G. E. Stelmach and J. Requin (Eds.), *Tutorials in motor behavior,* pp. 355-371. Amsterdam: Elsevier.

Greenwald, A. G. (1972). On doing two things at once: I. Time-sharing as a function of ideomotor compatibility. *Journal of Experimental Psychology, 100,* 52-57.

Greenwald, A., and Shulman, H. (1973). On doing two things at once: II. Elimination of the psychological refractory period. *Journal of Experimental Psychology, 101,* 70-76.

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74,*1464-1480.

Heuer, H. (1985). Intermanual interactions during simultaneous execution and programming of finger movements. *Journal of Motor Behavior, 17,* 335-354.

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

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

Johnston, J. C, and McCann, R. S. (forthcoming). On the locus of dual-task interference: Is there a bottleneck at the stimulus classification stage?

Kinsbourne, M. (1981). Single channel theory. In D. Holding (Ed.), *Human skills,* pp. 65-89. Chichester, U.K.: Wiley.

Kornblum, S. (1973). Sequential effects in choice reaction time: A tutorial review. In S. Kornblum (Ed.), *Attention mid Performance IV,* pp. 259-288. New York: Academic Press.

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

Logan, G. D. (1978). Attention in character classification tasks: Evidence for the automatic-ity of component stages, *Journal of Experimental Psychology: General, 107,* 32-63.

Logan, G. D., and Burkell, J. (1986). Dependence and independence in responding to double stimulation: A comparison of stop, change and dual-task paradigms. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 549-563.

Logan, G. D., and Zbrodoff, N. J. (1982). Constraints on strategy construction in a speeded discrimination task. *Journnl of Experimental Psychology: Human Perception and Performance, 8,* 502-520.

MacNamara, J., Krauthammer, M., and Bolgar, M. (1968). Language switching in bilinguals as a function of stimulus and response uncertainty, *Journal of Experimental Psychology, 78,* 208-213.

McLeod, P., and Posner, M. I. (1984). Privileged loops from percept to act. In H. Bouma and D. G. Bouwhuis, (Eds.), *Attention and Performance X,* pp. 55-66. London: Erlbaum.

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1423-1442.

Meuter, R. F. I., and Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40,* 25-40.

Meyer, D. E., and Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: 2. Accounts of psychological refractory-period phenomena. *Psychological Review, 104,* 749-791.

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

Monsell, S. (1996). Control of mental processes. In V. Bruce (Ed.), *Unsolved mysteries of the mind: Tutorial essays in cognition,* pp. 93-148. Hove, U.K.: Erlbaum and Taylor and Francis.

Norman, D. A., and Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, and D. Shapiro (Eds.), *Consciousness and self-regulation,* vol 4, pp. 1-18. New York: Plenum Press.

Osman, A., and Moore, C. (1993). The locus of dual-task interference: Psychological refractory effects on movement-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance, 19,*1292-1312.

Palmer, J. (1995). Attention in visual search: Distinguishing four causes of a set-size effect. *Current Directions in Psychological Science, 4,*118-123.

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 358-377.

Pashler, H. (1990). Do response modality effects support multiprocessor models of divided attention? *Journal of Experimental Psychology: Human Perception and Performance, 16,*826-840.

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

Pashler, H. (1997). *The psychology of attention.* Cambridge, MA: MIT Press.

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

Pashler, H., Carrier, M., and Hoffman, J. (1993). Saccadic eye movements and dual-task interference. *Quarterly Journal of Experimental Psychology, 46A,* 51-82.

Pashler, H., and Christian, C. (n.d.). Dual-Task interference and motor response production.

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

Pashler, H., and Johnston,]. C. (1998). Attentional limitations in dual-task performance. In H. Pashler (Ed.), *Attention,* pp. 155-189. Hove, U.K.: Psychology Press, Erlbaum, and Taylor and Francis.

Pashler, H., Luck, S. J., Hillyard, S. A., Mangun, G. R., and Gazzaniga, M. (1994). Sequential operation of disconnected cerebral hemispheres in split-brain patients. *Neuroreport, 5,* 2381-2384.

Rickard, T., and Pashler, H. (1998). A bottleneck in retrieval from a single cue. Talk presented at Thirty-ninth Annual Meeting of the Society for Psychonomics, Dallas, November.

Rogers, R. D., and Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124,* 207-231.

Ruthruff, E., Miller, J., and Lachmann, T. (1995). Does mental rotation require central mechanisms? *Journal of Experimental Psychology: Human Perception and Performance, 21,* 552-570.

Salthouse, T. A., Fristoe, N., McGuthry, K. E., and Hambrick, D. Z. (1998). Relation of task switching to speed, age, and fluid intelligence. *Psychology and Aging, 13,* 445-461.

Schvaneveldt, R. W., and Staudenmayer, H. (1970). Mental arithmetic and the uncertainty effect in choice reaction time. *Journal of Experimental Psychology, 85,*111-117.

Schouten, J. F., Kalsbeek, J. W. H., and Leopold, F. F. (1960). On the evaluation of perceptual and mental load. *Ergonomics, 5,* 251-260.

Seymour, P. H. (1973). Rule identity classification of name and shape stimuli. *Acta Psychologica, 37,*131-138.

Shaffer, L. H. (1965). Choice reaction with variable S-R mapping. *Journal of Experimental Psychology, 70,* 284-288.

Simon, J. R., Webste., J. B., and Craft, J. L. (1981). Effects of changing compatibility relationships on reaction time to the first stimulus in a double-stimulation task. *Journal of Motor Behavior,* 13, 301-312.

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

Suarez, J. N., Winstanley, G. and Griffiths, R. N. (1998). A reluctance-based cost distribution strategy for multiagent planning. *Applied Intelligence, 9,* 39-55.

Sudevan, P., and Taylor, D. A. (1987). The cueing and priming of cognitive operations. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 89-103.

Telford, C. W. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology, 14,*1-36.

Treisman, A., and Davies, A. (1973). Dividing attention to ear and eye. In S. Kornblum (Ed.), *Attention and Performance IV,* pp. 101-117. New York: Academic Press.

Van Galen, G. R, and ten Hoopen, G. (1976). Speech control and single channelness. *Acta Psychologica, 40,* 245-255.

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. (1967). Single-channel operation in the brain. *Acta Psychologica, 27,* 5-22.

Welford, A. T. (1980). The single-channel hypothesis. In A. T. Welford (Ed.), *Reaction time,* pp. 215-252. New York: Academic Press.