Journal of Experimental Psychology:
Human Perception and Performance
1994, Vol. 20, No. 2. 330-342

0096-1523/94/$3.00

Graded Capacity-Sharing in Dual-Task Interference?

Harold Pashler

Research suggests that dual-task interference is caused by a central bottleneck (together with response grouping and impaired preparation). The emphasis placed on the first response in these experiments, however, may have discouraged the sharing of processing resources between tasks. In the present experiment, instructions placed equal emphasis on 2 choice reaction-time tasks in which stimuli were presented simultaneously on 20% of the trials. In contrast to a graded trade-off of resources, a bottleneck predicts bimodality in the distribution of interresponse intervals for the 2 tasks, reflecting the 2 possible orders in which their respective central stages might be performed. Most subjects showed such a bimodality, along with other signs of a bottleneck; the remainder showed evidence of response grouping. The data suggest that the bottleneck is structural rather than strategic and make the graded sharing of resources less plausible.

People sometimes have difficulty performing two activities at the same time. One generally notices this only when one or both tasks are difficult, but in the laboratory very substantial dual-task delays occur with even the simplest tasks. The present article addresses a very basic question about the causes of this interference: Is it attributable to graded sharing of "mental capacity" or "resources" between the tasks or to a discrete processing bottleneck whereby certain operations in one task regularly delay operations in the other task?

The simplest experimental situation that demands concurrent performance of two tasks is the so-called psychological refractory period (PRP) paradigm. Here, two stimuli (S1 and S2) are presented, separated by some stimulus onset asynchrony (SOA), and the subject makes a separate speeded response to each (R1 and R2). Early experiments required manual responses to visual stimuli (e.g., Vince, 1949) and found a marked slowing of the second response (R2) as the SOA was decreased (the PRP effect). In some cases, the slope of the function relating R2 latency to SOA was about -1 for short SOAs (Welford, 1952), which amounts to saying an upper time limit was observed regarding the extent of task overlap.

The tasks used in these studies involve what most people would judge to be easy or even trivial sorts of activities, which makes the interference all the more remarkable. The interference also turns out to be quite robust. Variables that have been found to mitigate certain forms of dual-task interference (e.g., that caused by concurrent perceptual tasks or concurrent memory load effects) do not seem to reduce the PRP effect much. For example, whereas the earliest PRP studies tended to use two visual stimuli, PRP interference is not restricted to such cases. Davis (1959) found a PRP effect with a tone and a visual stimulus, for example, and many recent experiments have done likewise (see Pashler, 1993a, for a review). Nor is the effect restricted to tasks involving two manual responses. For example, classic PRP functions have been observed with manual and vocal response combinations (e.g., Pashler, 1989, 1990), manual and foot responses (Osman & Moore, 1990; Pashler & Christian, 1991), and even certain manual-oculomotor combinations (Pashler, Carrier, & Hoffman, 1993). Simultaneous visual-discrimination performance (with unspeeded responses) shows less interference when both stimuli are attributes of a single object (Duncan, 1984), but this is not true of the PRP effect (Fagot & Pashler, 1992). The PRP effect is also not readily eliminated with extensive practice (Gottsdanker & Stelmach, 1971; Johnston & Pashler, 1984), although much more investigation of this is needed before any definitive statement can be made. Finally, the PRP effect is not merely an artifact of the need to respond to two individual punctate stimuli isolated in time; PRP effects have been observed when one task was performed repeatedly before the second task appeared (Pashler & Johnston, 1991).

Causes of the PRP Effect

What causes people to have such trouble in carrying out what seem—computationally speaking—to be rather trivial tasks? A range of theories have been proposed since the PRP effect was first noted. Early suggestions that the effect might stem simply from temporal uncertainty about when S2 would arrive were rejected, although this factor sometimes contributes to a small degree (Bertelson, 1966; Pashler et al., 1993). A very natural idea about the PRP effect is that it is due to a "bottleneck," whereby some critical stages of each task occupy a common mechanism. According to this view,

---

This research was supported by National Institutes of Mental Health Grant 1-R29-MH45584-01. The author is very grateful to Alann Lopes for excellent technical assistance with this project, Allison Titus and Mike Lostetter for help in collecting the data, and James C. Johnston, Toby Mordkoff, Allen Osman and especially Jeff Lubin for helpful comments and discussion. Rich Carlson, Tom Carr, and Ray Klein made useful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Harold Pashler, Department of Psychology C-0109, University of California at San Diego, La Jolla, California 92093. Electronic mail may be sent to hpashler@ucsd.edu.

the critical stages in the second task cannot begin until the critical stages in the first task have been completed. This has the virtue of directly predicting the -1 slopes noted above. Different researchers suggested different accounts of what the critical stages or operations might be, however. Broadbent (1958) suggested a bottleneck in perceptual processing, whereas Welford (1952, 1980) postulated a bottleneck in the process of deciding what response should be made (what he termed the S-R "translation" process). Others suggested a bottleneck in actually initiating or carrying out motor responses (Keele, 1973).

Some early reviews favored Welford's response-selection bottleneck (e.g., Smith, 1967), but the evidence was not compelling, and the alternative idea of a response-production bottleneck continued to receive support (Logan & Burkell, 1986). In the intervening years, still further alternatives have been advanced. Of particular importance for the present article is the suggestion of graded sharing of capacity between tasks (Kahneman, 1973; McLeod, 1977). According to this view, mental resources can be divided between tasks in a graded fashion, with the efficiency of each task determined partly by the amount of resource(s) available to the task. The latter suggestion was motivated primarily by the observation that the first response (R1) as well as R2 is sometimes slowed in the dual-task situation, which can be readily explained if Task 1 as well as Task 2 suffers partial depletion of capacity. Many reviewers suggested that the idea of graded capacity-sharing (possibly involving multiple resources) provided the most reasonable picture of human information processing limitations (e.g., Broadbent, 1982; Wickens, 1983), and the value of the concept was simply assumed in studies designed to "measure spare mental capacity" (Posner & Boies, 1971).

Recent studies have examined the PRP effect using methods that appear to offer more critical tests of the various theories. One of these methods involves manipulating the duration of component stages of the second task (cf. Sternberg, 1969; Schweickert, 1978). Bottleneck models make very detailed and distinctive predictions for such experiments.

Figure 1 illustrates the time-course of dual-task performance according to the response-selection bottleneck model. Various predictions can be derived from this model. If an experimental factor is manipulated in such a way as to retard stages of the second task at or after the bottleneck (i.e., response selection or production), then this factor should slow the response correspondingly, whatever the SOA. For example, regardless of SOA, a 50-ms slowing of response selection in Task 2 should increase reaction times (RTs) by 50 ms (in both the dual-task and single-task conditions). On the other hand, if a stage in the second task located before the bottleneck (i.e., perceptual processing) is retarded, a counterintuitive result is predicted. In the single-task condition, and at very long SOAs, the full slowing should show up in the R2 latency. However, as the SOAs are reduced, completion of the second task becomes progressively more likely to have to wait on completion of the critical stages of the first task, rather than for the completion of the perceptual stages of the second task (the figure shows just such a short SOA). For that reason, the effect of the manipulation should be progressively reduced as SOA is shortened.

It can be seen, then, that different bottleneck models make distinctive predictions about effects of S2 factors: (a) additive effects of dual-task slowing (i.e., dual-task short SOA vs. long SOA; or dual- vs. single-task) and factors affecting stages of Task 2 beyond the bottleneck stage, and (b) underadditive interactions of SOA with factors affecting stages of Task 2 before the bottleneck should be observed.

A number of PRP experiments have been analyzed in more or less this way over the years. Both Keele (1973) and Schweickert (1978) analyzed previously published PRP experiments that compared simple and choice reaction times to the second stimulus (RT2s), looking for signs of processing "slack." Since it is unlikely that choice and simple RT differ merely in the duration of one particular single stage (Frith & Done, 1986; Sternberg, 1969), conclusions about dual-task interference derived from such

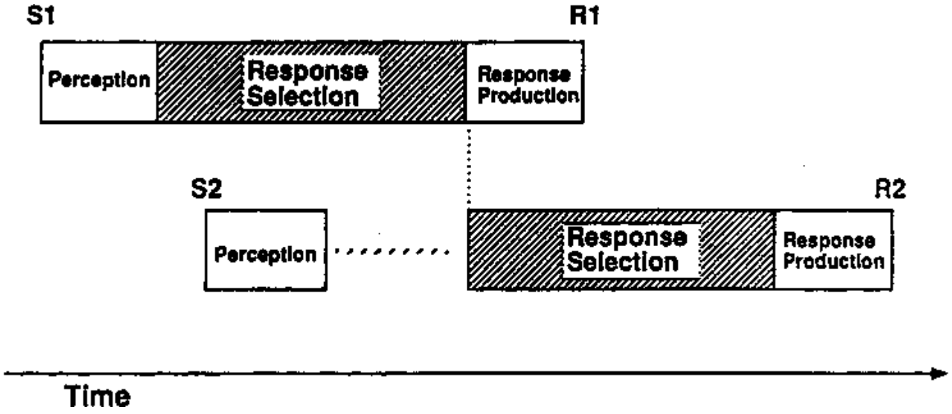


Figure 1. Response-selection bottleneck model: Response selection on Task 1 must be completed before response selection on Task 2 can begin. (S1 and S2 = stimulus that appears in Task 1 and Task 2, respectively; R1 and R2 = response time to S1 and S2, respectively.)

analyses are uncertain. In the past several years, a number of studies have used Task-2 factors chosen to affect the duration of specific stages of processing. The results have been quite consistent. When perceptual processing is slowed by reducing the intensity of a visual S2, the RT2 is increased less in the dual-task condition than in the single-task (Pashler, 1984), and in the dual-task condition, intensity effects are reduced as SOA is shortened (Pashler & Johnston, 1989). Similarly, display-size effects in a visual search second-task are underadditive with dual-task slowing (Pashler, 1984). These effects indicate that there is a bottleneck beyond the perceptual processing stage. On the other hand, manipulations that affect the duration of response selection generally have effects that are additive with dual-task slowing. For example, when the stimulus is repeated from trial to trial in a choice RT experiment, response selection is quicker (Pashler & Baylis, 1991); this effect is additive with dual-task slowing and with SOA (Pashler & Johnston, 1989). Similarly, McCann and Johnston (1992) found that S-R compatibility was additive with SOA.

The results support a response-selection bottleneck as shown in Figure 1. They also rule out the hypothesis that *producing* R2 is the first or only stage of processing that is delayed (as proposed by Keele, 1973, and Norman & Shallice, 1986). If no stage in Task 2 before response production is delayed, then R2 latencies are not affected by "cognitive" factors such as S-R compatibility at short SOAs.

There is another important but often neglected factor in dual-task interference, namely preparation. The operations comprising an arbitrary choice task will not occur at all unless the subject adopts the intention of carrying out the task, and they work more efficiently when the subject has ample opportunity to prepare, and little else to prepare at the same time. It appears that in a PRP situation, people typically prepare both tasks, and furthermore, that they prepare to carry out the tasks in a particular order. Changes in preparation seem to take place over a comparatively slow time scale (roughly speaking, several hundred milliseconds to a second).

These conclusions seem to be required by a number of observations, of which the following are three examples. First, when the order of stimuli is unknown to the subject, performance is impaired, most especially with pairs of manual responses (Pashler, 1990). Second, even when the SOA is very long (longer than the reaction time to the first stimulus, or RT1), RT2 is often slower than a single-task control (Welford, 1980). Third, performing the first task repeatedly before the stimulus for the second task appears slows responses to the second (but not the first) task, whereas bottleneck interference remains (Pashler & Johns-ton, 1991). In short, there is interference between the tasks both before and after the stimuli appear, but the interference is of a different sort. The most reasonable account of all of this would seem to be that (a) people have a limited "capacity" for preparing multiple tasks in advance (see Gottsdanker, 1980); (b) the degree of advance preparation affects the speed of all stages, including response selection; but (c) when one response selection operation begins, all others must wait. Pashler (1993b) presents this argument in more detail.

Further evidence for bottleneck models comes from analyzing the dependencies between R1 and R2 latencies. If processing required to produce R2 waits for completion of the major portion of Task 1 (where most of the variability in latency would be expected to originate), then most trial-to-trial variance in R1 latency should propagate into Task 2. Consistent with this, Welford (1967) noted positive correlations between RTs in the two tasks. Positive correlations of this sort might be consistent with other possibilities, however, such as a positive correlation between the degree to which each task was prepared in advance. Bottleneck models predict a much more specific pattern of dependency between R2 and R1 latency, interacting with SOA. In various analyses (e.g., Pashler, 1989), the trials were divided into quintiles, according to the relative speed of R1 within a condition. Figure 2 shows the mean R2 latency as a function of the relative R1 latency in a representative PRP experiment. At the long SOAs, there is little relationship (flat slope). As the SOA is reduced, R2 depends more positively upon R1 (upward tilt). Any bottleneck hypothesis predicts just such an interaction between SOA and R1-latency quintile. We have observed it repeatedly in PRP experiments with varying tasks, response modalities, and levels of practice (e.g., Fagot & Pashler, 1992; Pashler, 1989, 1991; Pashler & Johnston, 1991; Pashler & O'Brien, 1993).

Capacity-Sharing Revisited

The evidence just reviewed has been used to argue against capacity-sharing, but there is actually a glaring weakness in the argument. In almost all of the experiments described here, subjects were told to produce the first response as rapidly as possible. Suppose that graded sharing of capacity were a possible strategy available to the subject. Since an intermediate allocation of capacity would slow down the first task, subjects might well avoid it in order to comply with the instructions. Thus, they might allocate full capacity to the first task until it had been completed, thereby mimicking a bottleneck. (To make the capacity-sharing model fit the results reviewed above, one would have to postulate that the capacity is used only for response selection, rather than for any difficult parts of the task, as Kahneman [1973] proposed, and also that it can be rapidly reallocated once a task has been completed.)

There was one study among those listed above, however, in which subjects were not required to make the first response as rapidly as possible (Pashler & Johnston, 1989, Experiment 2). In that experiment, subjects were encouraged to produce the two responses at roughly the same time. This produced a strategy usually termed "grouping" (of the sort described by Borger, 1963, rather than Welford, 1980), in which RT1 was lengthened as SOA was lengthened. The findings of that study argued in favor of an analysis similar to that shown in Figure 2 (Panel C), in which the response-selection bottleneck still operated, but where the production of R1 was delayed until R2 had been selected. (The only difference was that the responses were produced in sequence, separated by

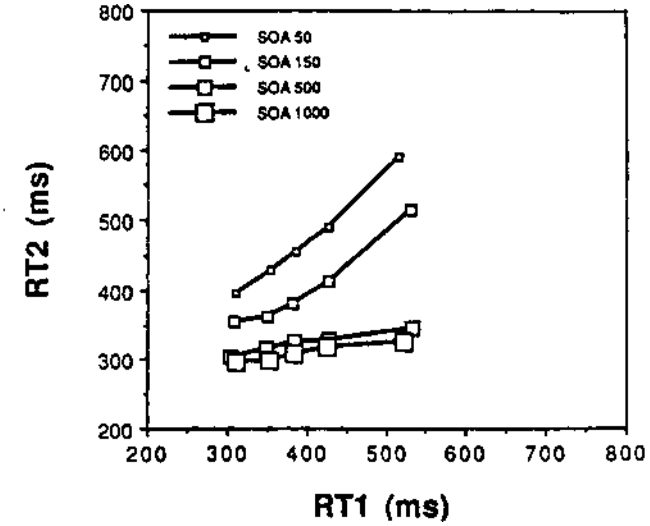


Figure 2. Second response time (RT2) as a function of speed of corresponding first response time (RT1). (For a given stimulus onset asynchrony [SOA], the five connected points represents the mean RT1 and RT2 for all those trials in which RT1 lay within quintile 1-5). From "Dual-Task Interference and the Cerebral Hemispheres," by H. Pashler and S. O'Brien, 1993, *Journal of Experimental Psychology: Human Perception and Performance, 19,* p. 328. Copyright 1993 by the American Psychological Association. Reprinted by permission.

a very short IRI with low variability.) The results would certainly seem to reduce the likelihood that subjects can allocate capacity arbitrarily among tasks: if a 50-50 allocation were possible, would it not have been a reasonable way to produce responses at roughly the same time? However, it could be argued that the instructions may somehow have discouraged this strategy—perhaps because it would produce such variable interresponse intervals (IRIs)—so the possibility of capacity sharing can still be salvaged despite these results.

In summary, there is fairly strong evidence that response selection typically constitutes a bottleneck when the instructions stress rapid production of the first response, and in at least one case where the instructions favored responding close together in time. However, we know little about what happens when people actually try to place equal emphasis on two tasks, without any encouragement of grouping. These are the conditions under which graded capacity-sharing would seem most likely to occur, and if under these conditions there were still no sign of capacity-sharing, this would seriously raise questions regarding its existence. The present experiments were undertaken to address this question.

The Present Approach

The basic idea was very simple. Subjects performed two tasks, involving a left-hand response to a tone and a right-hand response to a letter. The SOA between tone and letter was -1000 ms, -500 ms, 0 ms, 500 ms, and 1000 ms. The instructions did not mention grouping or response patterning; subjects were just told to place equal emphasis on each task. The 0-SOA condition is most critical for the analysis, but a wide range of SOAs were included because pilot work suggested that without them, response grouping emerged as the dominant strategy.

If the bottleneck model is true, then subjects ought to have only two choices about how they schedule response selections: They can select the response to the tone first, or they can select the response to the letter first. These are shown in Panels A and B of Figure 3. It can be assumed that when one response is selected first, it will be produced first. Given the results reported by Pashler and Johnston (1989), however, subjects might also choose to group responses (Panel C). In that experiment, subjects were instructed to respond to the two stimuli in a fixed order and not simultaneously, whereas with the present instructions, there is nothing to discourage simultaneous production of the responses, and therefore that is what is shown in the figure.

From these considerations it can be seen that the bottleneck model predicts that the resulting bivariate RT distribution should reflect a mixture of three types of trials: ungrouped tone-response-first trials, ungrouped letter-response-first trials, and grouped trials. (For grouped trials one could distinguish the two possible orders in which responses are selected; the distinction will not matter if responses are simultaneous.) A priori, there is no way of knowing whether grouping is likely to be confined to individual subjects, or whether some subjects will engage in all three of these strategies. Our previous results indicate that the variance of response selection is a major contributor to the variance of observed RTs, whereas the variance of IRIs produced by a grouping strategy is extremely small (Pashler & Johnston, 1989). On this basis, various predictions about observed distributions can be derived, and they will be assessed below. Equally important, the bottleneck

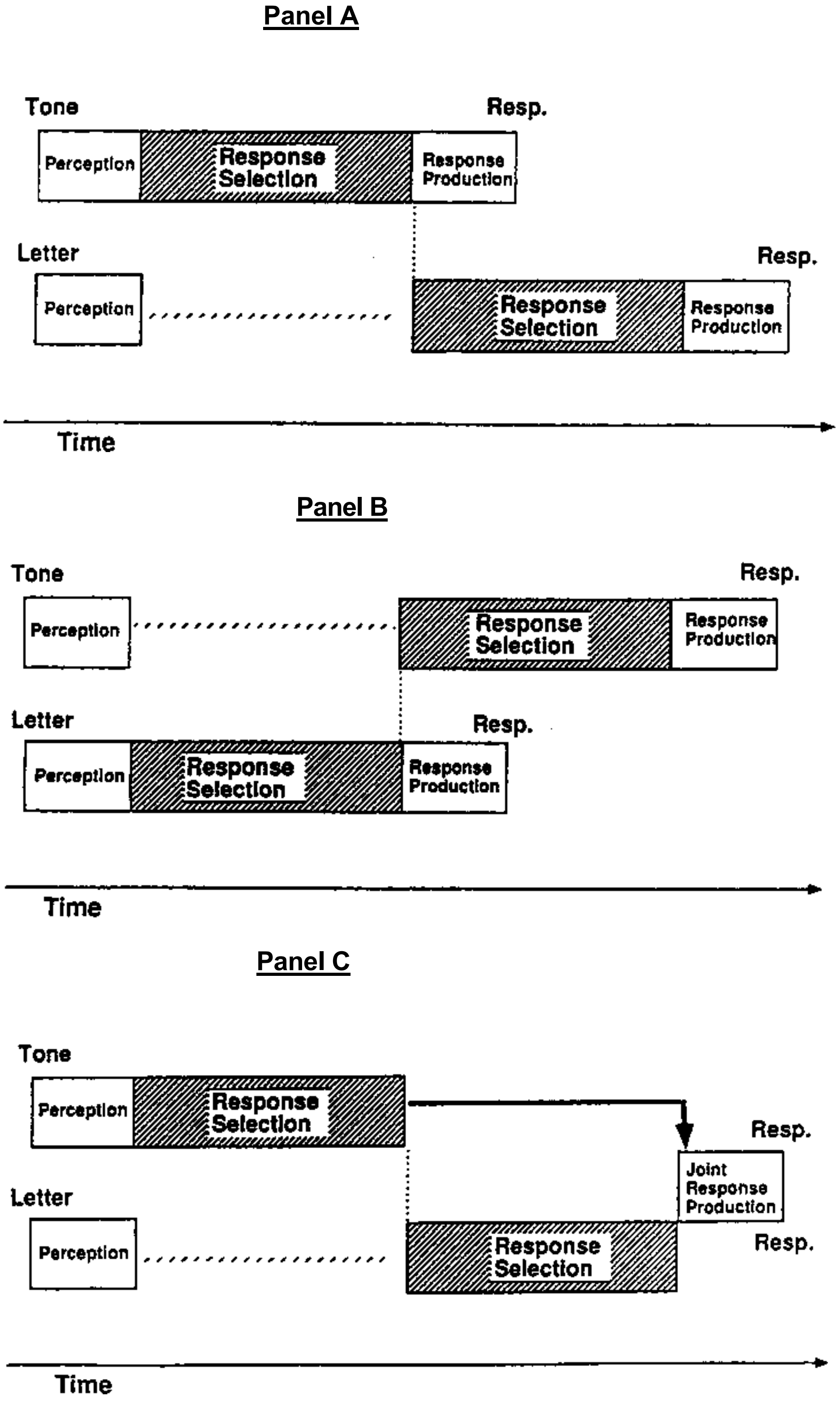


Figure 3. The response-selection bottleneck predicts that responses to simultaneous tone and letter will reflect three possible underlying sequences: select tone-response first (Panel A); select letter-response first (Panel B); select responses in series (one of two possible orders is shown), and then produce responses as a couplet ("response grouping"; Panel C).

model makes a clear prediction about what should *not* be observed: One should not find a broad smooth distribution of IRIs extending both positively and negatively about zero.

Capacity-Sharing Predictions

On the other hand, if people can split their capacity among the tasks arbitrarily, our instructions would seem to encourage a more or less even split about as effectively as could be done. (Of course, feedback designed specifically to discourage variability in IRIs would doubtless be more effective, but it would also encourage grouping; our impression is that it would produce grouping exclusively.) So what should be expected on trials where a roughly even allocation of capacity is adopted? Variability in the completion times for each task should produce a broad distribution of IRIs, presumably centered around zero. There is no reason to expect any bimodality in the IRIs. Predictions about the correlation of the RTs for the two tasks are less clearcut because three factors could affect this correlation. First, trial-to-trial variability in the proportion of capacity allocated to each task on different trials should tend to make the correlation negative. On the other hand, trial-to-trial variability in the total amount of capacity available, and rapid reallocation after completion of the first task, should both tend to make the correlation positive. The net result cannot be anticipated. What about grouping? The occurrence of this strategy on some trials would be consistent with the underlying potential for sharing capacity; therefore, grouping cannot discriminate between the models.

Method

Subjects

Twenty-four students from the University of California at San Diego participated in partial fulfillment of a course requirement.

Apparatus and Stimuli

The stimuli were presented on NEC Multisync 2 or 2a monitors, controlled by IBM-PC microcomputers (equipped with a Paradise VGAPlus graphics card, providing a display resolution of 640 X 200 pixels). The stimuli consisted of one tone and one letter on each trial. The letter was either a *Q, T,* or *V,* presented centrally in white against a black background. The letters were .6 cm wide and 1.3 cm high (.6° x 1.2°, based on a typical viewing distance of 60 cm). The tones were either 300 Hz or 900 Hz in frequency, presented through the speakers on the monitor. Subjects responded by depressing keys on the microcomputer keyboard.

Design

The experiment was divided into 12 blocks of 40 trials each. There were five different SOAs (-1000, -500, 0, 500, and 1000 ms), defined as the time from the onset of the tone to the onset of the letter. The letter and tone presented on each trial were selected randomly without constraint. The different SOA conditions were presented to the subjects in random order, with 96 trials at each SOA per subject.

Procedure

Subjects were given written instructions describing the tasks. The instructions described the two tasks, and stated that the subject should "place about equal emphasis on each task." The instructions also stated that subjects should not "put more effort unto responding quickly to one task or the other. They are equally important, whichever stimulus comes first." Prior to the collection of data, subjects completed one practice block containing 40 trials.

Nothing was done to encourage or discourage any particular pattern of responding, with one exception: The experimenter interceded with a few subjects whose strategy during the practice trials was to produce the responses in a fixed order, even when that order was opposite to that of the stimulus presentations at the longest SOA. In these cases, the experimenter pointed out to the subjects that they seemed to be placing more emphasis on responding quickly in whichever task they were habitually performing first. In no case was anything said to suggest that the two responses should or should not be produced at approximately the same time, however.

Each trial began with the presentation of the cross as a warning signal and fixation point. It remained present for 1 s, followed by a 500-ms offset. At this point, the tone and letter appeared in whatever sequence was dictated by the SOA. The letter remained present until both responses were detected. On the other hand, the tone lasted for 150 ms (tones that remained on until response were found to be somewhat annoying). Subjects responded to a high or low tone by pressing the *A* or *Z* key, using the middle or index finger of their left hand, respectively. They responded to the letter *Q, T,* or *V* by pressing the comma, period, or slash keys on the keyboard, using their right index, middle, or fourth finger, respectively. Subjects kept their fingers rested on the keys throughout the block.

Normally, the intertrial interval between completion of responses on a given trial and onset of the fixation point for the next trial was 1.5 s. If an error was made a visual warning appeared for a half second, followed by an extra 800 ms inserted in the intertrial interval to allow "recovery." At the end of each block the subject rested and received feedback regarding the percentage of correct response and average RT for each of the two tasks, for that block and each of the preceding blocks.

Results and Discussion

Basic Results

RTs below 150 ms or in excess of 1800ms were considered deviant and excluded. Figure 4 shows mean RTs for each task as a function of SOA. Not surprisingly, the effect of SOA was significant in both the tone RTs, $F(4, 92) = 183.5$, $p < .001$, and the letter RTs, $F(4,92) = 165.2$, $p < .001$. RTs are slowest at 0 SOA. For both tasks, the stimulus was responded to faster when it occurred at the end of a long SOA, namely tone-first-1000-ms SOA for the letter task, and letter-first-1000-ms SOA for the tone task. This pattern has been noted previously (Pashler, 1990) when the order of two stimuli is unknown to the subject (and especially when both responses are manual, as here). Presumably, the critical factor here is that after the first stimulus appears, the subject can infer which task is coming next. In line with the account of preparation noted above, the speedup in the second task at long

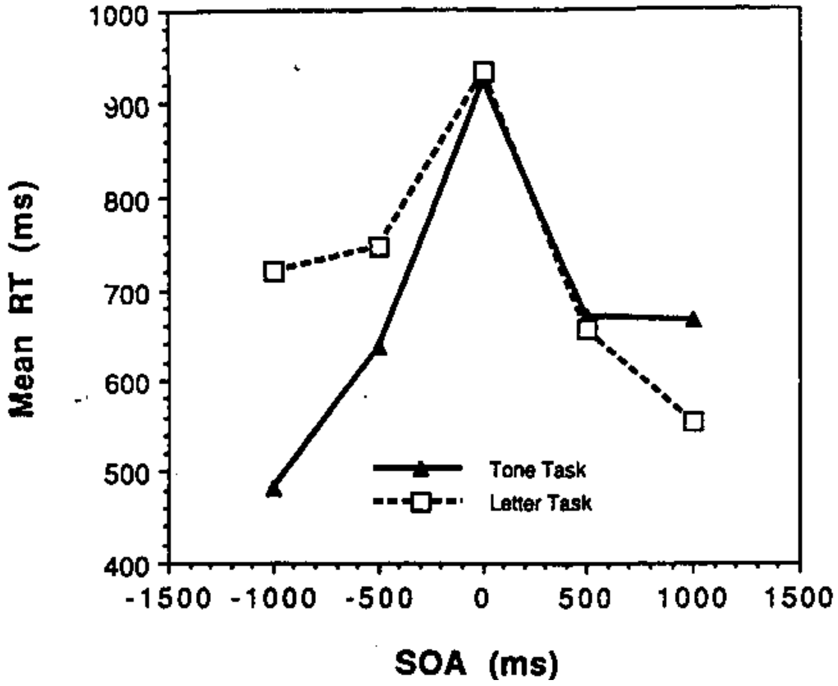


Figure 4. Mean response times (RTs) for each task as a function of stimulus onset asynchrony (SOA), measured from tone to letter.

SOAs can be attributed to a (relatively slow) change in preparation that occurs after the first task and that favors the second task. Indeed, it is hard to think of any other account of this pattern.

For the tone task, the error rates were 3.5%, 4.4%, 3.6%, 1.6%, and .8%, for SOAs of letter-first-1000 ms, letter-first-500 ms, 0 ms, tone-first-500 ms and tone-first-1000 ms, respectively. The difference was significant, $F(4, 92) = 16.4$, $p < .001$. The corresponding error rates for the letter task were 1.5%, 1.0%, 2.5%, 4.3%, and 3.2%, respectively, $F(4,92) = 8.0$, $p < .001$. Thus, subjects are slightly but significantly more accurate when responding to a stimulus that occurs first. The preparation factor noted above seems likely to be responsible for these effects as well.

R1-by-R2 Scatterplots

The main purpose of the experiment was to determine what happens at the 0-ms SOA when subjects are told to allocate equal capacity to the two tasks. According to the bottleneck hypothesis, RTs should reflect a mixture of three types of trials: select tone-response first, select letter response first, and response grouping (Figure 3). Figure 5 shows four scatterplots of individual RT(tone)-RT(letter) pairs that were judged fairly representative of the 24 scatterplots (each panel contains a dot for every acceptable response pair the subject made at 0 SOA). The panels on the left show results from two subjects in which almost all the points lie very tightly clustered on the diagonal. For these subjects, RT(tone) and RT(letter) are almost always equal, and the correlation of RT(tone) and RT(letter) was extremely high. By an informal assessment, 6 of the 24 subjects appeared to be of this type. The panels on the right show examples of what appeared to be a different and more common kind of scatterplot. Here there are almost no trials on which RT(tone) and RT(letter) are close in value. Rather, RT(tone) is either substantially greater than RT(letter) or it is substantially less. About 17 subjects appeared to be of this type, some with a preponderance of trials of one order or the other. Only 1 subject seemed to show features of both patterns.

IRIs

As a step toward analyzing this typology more quantitatively, the distribution of IRIs was computed for each subject. Figure 6 shows four additional subjects' IRI distributions (two of each of the types noted). The first set of subjects show IRI distributions that are basically a spike near zero while the second set show IRI distributions that are broadly smeared over a wide range of positive and negative values, but which conspicuously exclude zero. To demonstrate that subjects really do clump into two groups, a statistic was computed for each subject: the proportion of trials on which the IRI fell within the range -75 to +75 ms. Figure 7 shows a histogram of the number of subjects for whom this statistic falls at each range from (0-.10) to (.90-1.0). Seventeen subjects' proportions lay below .30 (mostly very near zero), 6 subjects' proportions lay above .70, and 1 (apparently intermediate) subject showed an intermediate value. The bimodal distribution of subjects is plain.

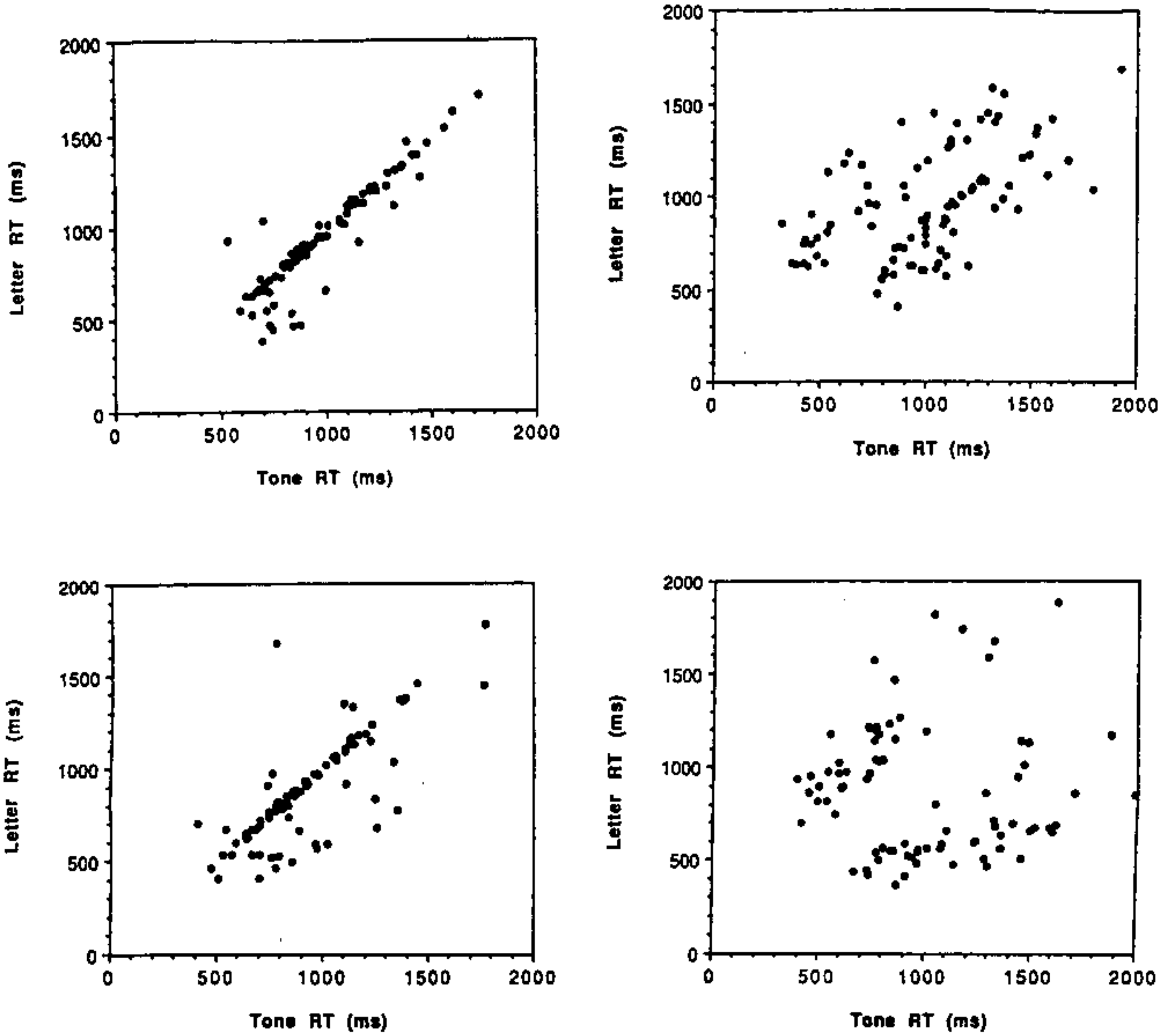


Figure 5. Scatterplots for each of four subjects showing reaction time (RT) to tone (Tone RT) and RT to letter (Letter RT) for each 0-SOA trial with correct responses within bounds: two "spike subjects" (top left and bottom left); two "double ridge" subjects (top right and bottom right).

The main result, then, is that subjects fall into two types. These will be labeled spike subjects and double-ridge subjects, with reference to the shape of their IRI distributions. The spike subjects show the behavior expected from response grouping (Figure 3C). The double-ridge subjects exhibit the mixture of two processing orders anticipated on the basis of the bottleneck model (Figures 3A and 3B).

It is interesting that when the stimuli were asynchronous, the double-ridge subjects almost always responded to them in the order they were presented. (As noted above, we explicitly discouraged subjects from responding in a single fixed order throughout, but the vast majority of subjects needed no such discouragement.) Among acceptable accurate responses, the average number of reversals per subject was zero at the letter-first-1000 and tone-first-1000 SOAs, 1.35 at the letter-first-500 SOA, and .47 at the tone-first-500 SOA (out of a possible 96 trials).

Characteristics of Spike and Double-Ridge Subjects

Given how they were selected, it is not surprising that the average IRI variance was lower for the spike subjects (18,268) than for the double-ridge subjects (103,777). If double-ridge subjects" latencies reflect a mixture of two different processing orders (select tone-response first vs. select letter-response first) as suggested here, then the correlation between RT(tone) and RT(letter) might be quite weak in the

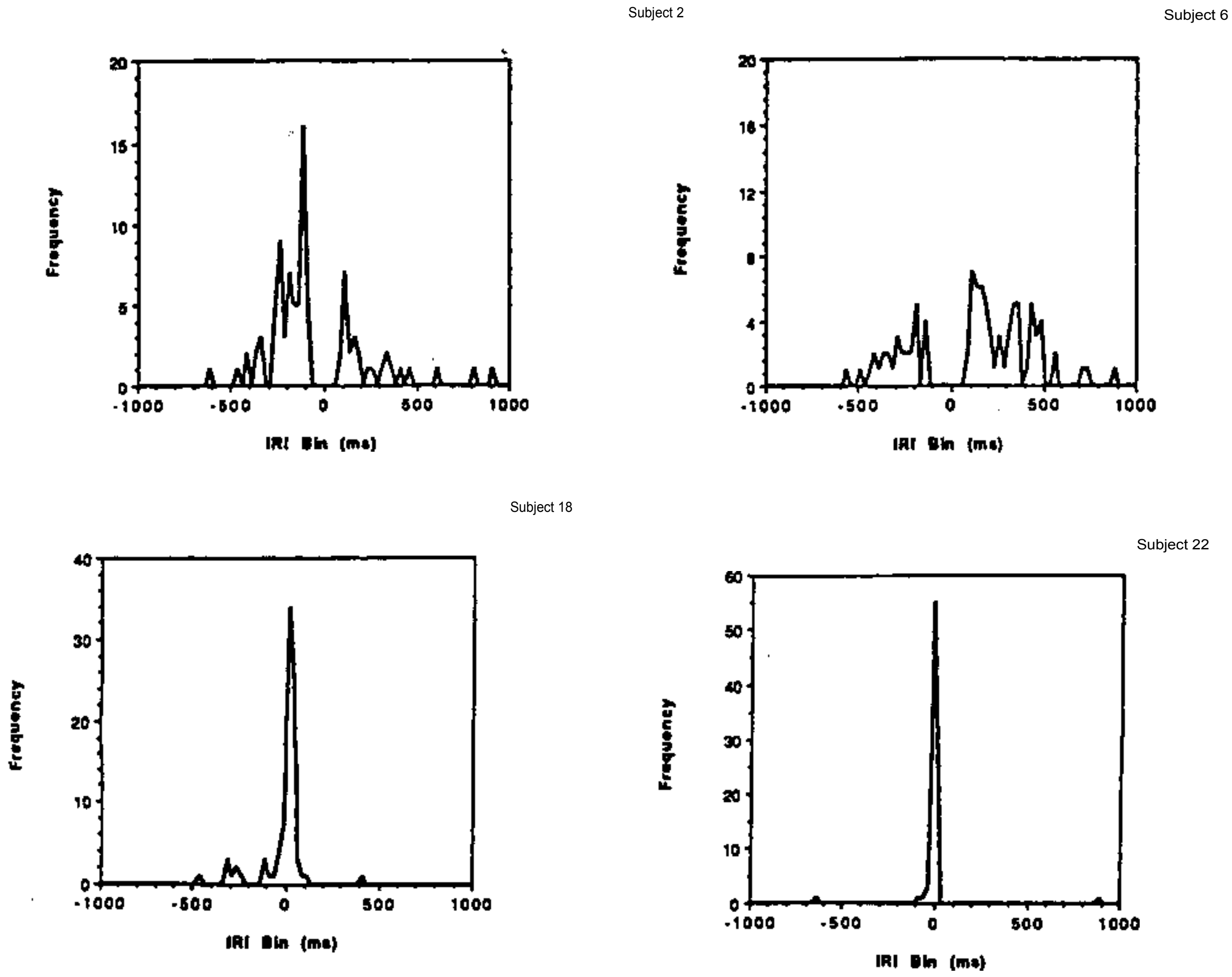


Figure 6. Distribution of interresponse intervals (IRIs) for four additional subjects, 0-SOA (stimulus onset asynchrony) condition. (Bin = IRI range.)

0 SOA trials generally. This is because although the bottleneck induces a positive correlation (if B waits for A, then random variation in A propagates onto B), the inclusion of a mixture of the two processing orders superimposes some negative correlation on top of that positive correlation. Thus, one would expect that when the correlation is computed on just those trials on which responses came out in a particular order, then there should be the same high correlation generally observed in PRP experiments. To examine this question, for each subject the correlation of RT(tone) and RT(letter) was computed for all the 0 SOA trials and this was averaged across subjects. The mean was .362. When the analysis was restricted to just those trials where the letter response came first, the correlation was .766, and when it was restricted to trials where the tone response came first, it was .795. Whereas bifurcating a distribution might be expected to increase the correlation on any account, it is encouraging to find that each of the resulting correlations are in the range expected from previous PRP experiments, in which the order of responding was controlled by instructions (e.g., Pashler & Johnston, 1989).

Most of the double-ridge subjects used both orders plentifully on 0 SOA trials. What determines the order in which they schedule the two tasks on a given trial? One potential factor is the SOA of the previous trial. If on a given trial the letter preceded the tone by 1000 ms, for example, that virtually insures that the letter response would have been produced first on that trial. Subjects might tend to respond to the letter first on the next trial (where the SOA was 0), if they tended to repeat the same task order. Alternatively, one might suppose that whichever individual task was done most recently would tend to be favored, thus producing the opposite

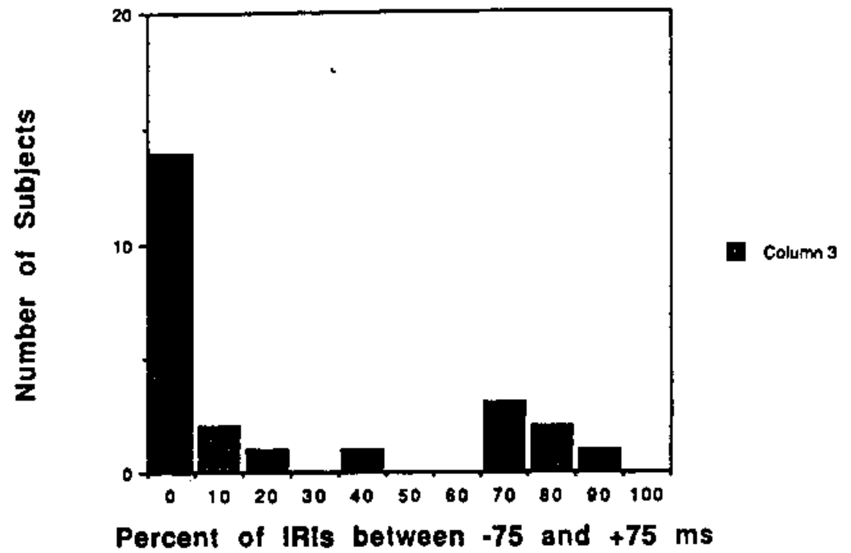


Figure 7. Histogram showing proportion of interresponse intervals (IRIs) lying within interval (-75 ms to +75 ms) for trials at zero stimulus onset asynchrony (SOA) for "double-ridge" subjects only.

sequential dependency. In fact, averaged across subjects, the tone response occurred first on 68% of 0-SOA trials that were preceded by a tone-first- 1000-SOA trial, compared to 45% of 0-SOA trials that were preceded by a letter-first-1000-SOA trial. This difference was significant, $F(1, 16) = 135.7$, $p < .001$. It shows that after an extreme SOA trial that "forces" subjects to respond in a particular order, they tend to respond in the same order to simultaneous stimulation on the next trial.

One might also imagine that when a letter is repeated on two successive trials (making the letter task "easier"), the response to the letter would come out first on the second of the two trials. For each of the double-ridged subjects, the proportion of trials on which the tone response came first was computed, depending on whether the letter was the same as on the previous trial. The averages were .543 and .554, respectively, which was not a significant difference ($F < 1$). There is no sign that response selection difficulty determines the order of queuing. Together, these results are certainly consistent with the notion that it is a state of task preparation attained prior to the stimulus that determines the task order at 0 SOA (which we suspect to be the case), rather than the order being determined by the exigencies of processing arising after the stimuli appear. However, these results do not necessarily force one to accept this conclusion.

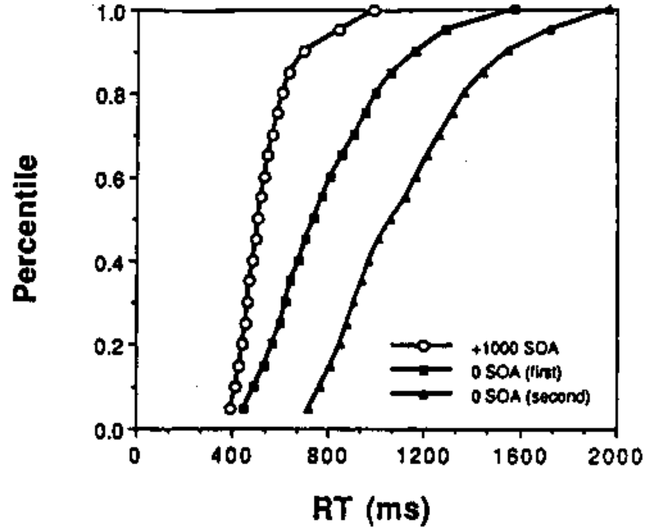
Response Latencies in Double-Ridge Subjects

The double-ridge subjects show obvious bimodality in their IRIs (see Figure 6). According to the bottleneck model, the positive and negative IRIs reflect trials on which the tone task was performed first or second, respectively. If *all* slowing were attributable to bottleneck-induced delays, then when a given task is performed first at 0 SOA, it should be responded to as quickly as in a long SOA condition, where the stimulus for this task came first. (Note that the two conditions being compared do not differ in terms of temporal relationship of the stimulus to the warning interval, making this comparison reasonable.) By this reasoning, then, one might expect that RTs for the tone task when it is performed first at 0 SOA would be the same as the RTs for the tone task with a 1000-ms SOA. In fact, the mean RTs in these two conditions were 758 and 654 ms, respectively. Thus, tone-task responses that precede letter responses are significantly slower when the letter is present along with the tone, compared to when the tone comes first and the letter comes well after. Of course, this slowing is a great deal less than the slowing observed with tone responses that are produced second at 0 SOA. There, the mean latency was 1067 ms. Exactly the same pattern was observed with the letter responses: RTs for the letter task were slower when the letter response was produced first in the 0 SOA condition (774 ms) compared to when the letter came 1000 ms before the tone (697 ms); however, responses for the letter task produced second at 0 SOA were a great deal slower (1055 ms).

At 0 SOA, why should the task that is performed first be performed more slowly, given that this task is presumed to receive priority treatment at the bottleneck? One would expect it to be just as fast. Recall, however, that unlike the traditional PRP situation, this paradigm required subjects to deal with stimuli that can appear in either order. When two stimuli arrive simultaneously, the process of determining which one is to be dealt with first may itself produce occasional delays. It might even be the case that subjects sometimes start processing one task, and then abort it and switch to the other task. Pashler (1990) specifically examined the costs of stimulus order uncertainty by comparing fixed and variable orders; the latter condition involved substantial costs for tasks with two manual responses, as in the present experiment (the costs were smaller for tasks with one manual response and one vocal response). It can be assumed, therefore, that these variable-order costs are contributing to 0-SOA slowing in the present situation as well.

One might suspect that such "supervisory costs" would introduce occasional (rather than constant) delays. If so, then the fastest responses should be less affected, since these will tend to happen on trials where the intermittent disturbances were minimal. Therefore, the cumulative distribution functions (CDFs) fir the RTs of these conditions should increasingly diverge. Figure 8 shows the Vincentized CDFs for these conditions (nongrouping subjects only), analyzed with 3-s cutoffs (along with those for the second-produced response at 0-SOA). The expected divergence is obviously present in both tasks, with much less cost associated with the 0-SOA condition in the fastest RTs compared to the slowest RTs. On the other hand, the slowing of the second-produced response is present throughout the distribution. The main implication of this figure, though, is that the 0-SOA slowing of the first-produced response may well reflect fairly intermittent events.

It may seem discouraging that a number of interlocking factors in even these relatively simple dual-task situations must be posited: the bottleneck, response grouping, preparation variability, and intermittent scheduling difficulties. It should be noted, though, that the last three of these factors seem to play only minor roles in the traditional PRP experi-


Responses to Tone

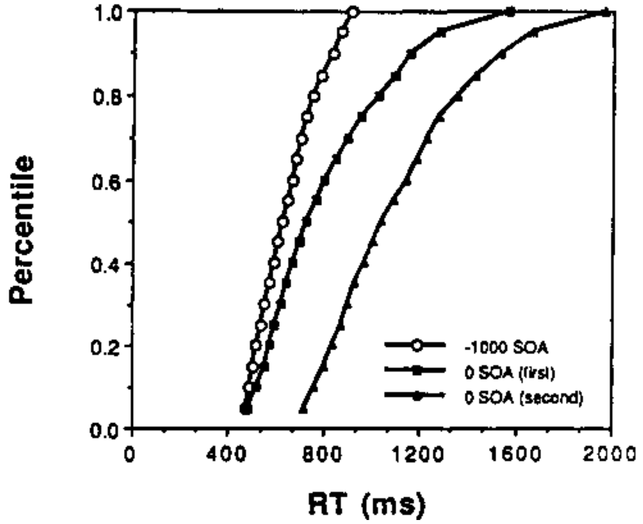

Responses to Letter

Figure 8. Vincentized cumulative response time (RT) density functions for each task, comparing three conditions: task stimulus preceded other stimulus by 1000 ms; response was produced first (stimulus onset asynchrony [SOA] = 0); and response was produced second (SOA = 0).

ment with short SOAs and instructions stressing rapid first-task responses. It is not really surprising that some additional factors—in fact, these particular factors—should emerge when the subject is faced with stimuli that can occur in either order and he or she is left to his own devices in choosing a response pattern. The contention of the present paper is not that every aspect of this situation can be fully unraveled at this point, but only that the basic causes of the interference can be discerned.[1]

General Discussion

Conclusions

The results strongly favor a bottleneck model. A bottleneck requires that critical processing stages are carried out in one order or the other (but not simultaneously in both tasks). If responses are produced as soon as they are selected the model predicts a bimodal pattern in the distribution of IRIs, in fact, such a pattern was observed for 16 of the 24 subjects. All of the remaining subjects except one showed strong evidence of response grouping, with extremely tight clustering of the IRIs about zero. The occurrence of this pattern was anticipated on the basis of previous research (Borger, 1963; Pashler & Johnston, 1989). Response grouping is entirely consistent with the existence of a fundamental response selection bottleneck (Figure 3C).

The only way in which the present results diverge from the predictions made from a pure bottleneck model is that even the responses produced first at the 0-SOA condition show some slowing compared to a very long SOA condition. Inspection of the RT distributions suggests that the amount of slowing is quite variable from trial to trial, which is consistent with the idea that it is caused by difficulty in coordinating the order of processing.

Limitations

The present results demonstrate that when subjects were told to place equal emphasis on both tasks, their behavior reflects a mixture of the underlying processes postulated by the response selection bottleneck account (modulated, in some cases, by response grouping). To put it simply, there is no sign that people can just choose to do the two tasks at the same time but more slowly. Rather, it appears that the critical stages must be carried out in one order or the other, and certain subjects then elect to group their responses. Thus, the results support the bottleneck account and question the reality of capacity-sharing.

Dual-task interference is a complicated business, however, and it would be unwise to overstate the conclusions. Thus, some limitations should be plainly acknowledged.

First, it is always possible that different results would be obtained with other kinds of tasks, other combinations of input and output modality, and so on. The fact that some signs of bottleneck have been obtained with manual-vocal combinations (Pashler, 1991) does not mean that the universality of the response selection bottleneck can be assumed. (Indeed, Pashler et al., 1993, showed that certain eye movements clearly seem to work independently of the bottleneck.) Furthermore, it has been noted that lack of advance knowledge of stimulus order produces interference with manual-manual combinations; for this reason, analyses such as those performed here—which use combinations of response modalities less subject to these preparation-related difficulties— would be of particular interest.

A second (and more fundamental) limitation is that results like the present ones can never be said to rule out the possibility that people can allocate capacity among different tasks in a graded fashion. Like any claim that *X* can sometimes happen, this possibility is not falsified by individual cases in which *X* did not happen. The case against *X* is strengthened to the degree one has looked in the circumstances in which *X* is most likely to happen. All that is claimed here is that this was a very natural place to expect smoothly varying capacity allocation to occur, if it ever does.

There are various ways in which the hypothesis of graded capacity-sharing might be pursued further. One could, for example, deliberately manipulate the speed of one of the two responses using something like a deadline method (Reed, 1976). According to the bottleneck model, certain temporal configurations of R1 and R2 should be extremely difficult, and perhaps impossible, for subjects to produce (at least with any reasonable level of accuracy). If subjects can really allocate capacity in a graded way, on the other hand, they ought to have a good deal of flexibility. One might also use a deadline to induce the subject to slow the first response a little beyond its usual value; according to a capacity-sharing model, taking a little more time for Task 1 ought to be achievable by giving it a little less capacity. This would not be likely to delay the second task. According to a bottleneck model, on the other hand, it would seem likely that subjects could comply by slowing the execution of Task 1 (perhaps increasing its accuracy), and thereby delaying critical processing in the second task (and RT2) correspondingly. So very fine-grained speed-accuracy experiments might provide an even more analytic way to look at dual-task interference.

A third limitation on the present results is that capacity-sharing may be an appropriate framework for understanding other kinds of interference besides the bottleneck interference associated with response selection. For example, interference due to simultaneous perceptual processes working in the same modality appears to arise quite independently of the response-selection bottleneck (Pashler, 1989). This kind of interference may fit the metaphor of capacity-sharing more closely than do the more central processes that underlie the PRP effect. Further detailed comparisons of different kinds of dual-task interference should therefore prove most illuminating.

---

[1] If this sort of situation is revealed to be complex, then commonly studied dual-task experiments in which subjects perform tasks complicated such as solving anagrams or continuous visual-manual tracking could well be so complex as to defy meaningful analysis at the present time.

References

Bertelson, P. (1966). Central intermittency twenty years later. *Quarterly Journal of Experimental Psychology, 18,* 153-163. Borger, R. (1963). The refractory period and serial choice-reactions. *Quarterly Journal of Experimental Psychology, 15,* 1-12.

Broadbent, D. E. (1958). *Perception and communication.* Elmsford, NY: Pergamon Press.

Broadbent, D. E. (1982). Task combination and the selective intake of information. *Acta Psychologica, 50,* 253-290.

Carrier, M., & Pashler, H. (1993). *The attention demands of memory retrieval.* Manuscript submitted for publication.

Davis, R. (1959). The role of "attention" in the psychological refractory period. *Quarterly Journal of Experimental Psychology, 11,* 211-220.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113,* 501-517.

Fagot, C., & Pashler, H. (1992). Making two responses to a single object: Exploring the central attentional bottleneck. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1058-1079.

Frith, C. D., & Done, D. J. (1986). Routes to action in reaction time tasks. *Psychological Research, 48,* 169-177.

Gottsdanker, R. (1980). The ubiquitous role of preparation. In G. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior* (pp. 315-371). Amsterdam: North-Holland.

Gottsdanker, R., & Stelmach, G. E. (1971). The persistence of psychological refractoriness. *Journal of Motor Behavior, 3,* 301-312.

Johnston, J. C., & Pashler, H. E. (1984, November). *Can two decision processes operate simultaneously?* Paper presented at the meeting of the Psychonomics Society, Austin, TX.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Keele, S. W. (1973). *Attention and human performance.* Pacific Palisades, CA: Goodyear.

Logan, G. D., & 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.

McCann, R., & Johnston, J. C. (1992). Locus of the single-channel bottleneck in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 471-485.

McLeod, P. (1977). Parallel processing and the psychological refractory period. *Acta Psychologica, 41,* 381-391.

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

Osman, A., & Moore, C. (1990, October). *Where is attention limited: The effects of dual-task interference on the lateralized readiness potential.* Paper presented at the 30th meeting of the Society for Psychophysiological Research, Boston, MA.

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. (1989). Dissociations and dependencies between speed and accuracy: Evidence for a two-component theory of divided attention in simple tasks. *Cognitive Psychology, 21,* 469-514.

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. (1991). Shifting visual attention and selecting motor responses: Distinct attentional mechanisms. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 1023-1040.

Pashler, H. (1993a). Dual-task interference and elementary mental mechanisms. In D. Meyer & S. Komblum (Eds.), *Attention and performance* (Vol. 14, pp. 245-264). Cambridge, MA: MIT Press.

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

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

Pashler, H., & Christian, C. (1991). *Dual-task interference and response execution.* Manuscript submitted for publication.

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

Pashler, H., & Johnston, J. C. (1991). *Repeated task performance and dual-task interference: Chronometric studies.* Manuscript submitted for publication.

Pashler, H., & O'Brien, S. (1993). Dual-task interference and the cerebral hemispheres. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 315-330.

Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review, 78,* 391-408.

Reed, A. V. (1976). The time course of recognition in human memory. *Memory and Cognition, 4,* 16-30.

Schweickert, R. (1978). A critical path generalization of the additive factor method: Analysis of a Stroop task. *Journal of Mathematical Psychology, 18,* 105-139.

Smith, M. C. (1967). Theories of the psychological refractory period. *Psychological Bulletin, 67,* 202-213.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Roster (Ed.), *Attention and performance* (Vol. 2, pp. 276-315). Amsterdam: North-Holland.

Vince, M. (1949). Rapid response sequences and the psychological refractory period. *British Journal of Psychology, 40,* 23-40.

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). San Diego, CA: Academic Press.

Wickens, C. D. (1983). Processing resources in attention, dual task performance, and workload assessment. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63-101). San Diego, CA: Academic Press.

Received August 3, 1992
Revision received June 15, 1993
Accepted June 16, 1993 •