Bimanual response grouping in dual-task paradigms

Jeff Miller  
University of Otago, Dunedin, New Zealand

Rolf Ulrich  
University of Tübingen, Tübingen, Germany

In three experiments we measured response time (RT) and peak force (PF) to investigate the grouping of left- and right-hand key press responses in a dual-task paradigm involving two independent go/no-go tasks. Within each task, a go stimulus within one of two modalities (i.e., visual versus auditory) required a response by one hand. In Experiment 1 with simultaneous go stimuli in the two tasks, responses appeared to be grouped in approximately 75–80% of trials, compared with nearly 100% grouping in a single-task condition requiring bimanual responses to the onset of any stimulus in either modality. In Experiment 2 with stimulus onset asynchronies (SOAs) of 0–400 ms between the two go stimuli, response grouping clearly declined as SOA increased, although some grouping was still evident even at the longest SOA. The same pattern was observed in Experiment 3 with the same range of SOAs but unpredictable stimulus order, suggesting that grouping is not strongly dependent on prior knowledge of the likely response order. These results emphasize the pervasiveness of response grouping in bimanual dual-task RT paradigms and provide useful clues as to its nature.

In recent years, the psychological refractory period (PRP) paradigm has been one of the main tools used to investigate the cognitive limits that emerge when people are asked to perform two tasks simultaneously. In the standard PRP paradigm, participants are presented with two stimuli ($S_1$ and $S_2$) temporally separated by a small stimulus onset asynchrony (SOA). The participant is required to respond separately to each stimulus ($R_1$ and $R_2$), and the reaction time (RT) associated with each task is measured ($RT_1$ and $RT_2$). The standard findings are that $RT_2$ is substantially increased when the two stimuli occur in close temporal proximity and that this increase lessens as the time between the two stimuli increases.

It is commonly suggested that participants in PRP tasks may sometimes engage in the strategy of response grouping, especially in PRP paradigms requiring participants to respond manually in both tasks (e.g., Borger, 1963; Pashler, 1994a; Pashler & Johnston, 1989). A participant using this strategy would select $R_1$ but then hold it in...
waiting until $R_2$ was also ready to be initiated, presumably because it is easier to emit two responses simultaneously than to emit them in rapid succession.

The possibility of response grouping is theoretically important, because the presence of grouping can change the predictions of models for the PRP task (e.g., Ulrich & Miller, 2007). The well-known response-selection bottleneck model (RSBM; e.g., Pashler, 1984; Welford, 1952), for example, predicts that $RT_1$ should not depend on either SOA or Task 2 characteristics in the absence of response grouping. Observed dependencies of $RT_1$ on these manipulations can, however, be reconciled with this model under the assumption that participants sometimes group their responses (e.g., Hommel, 1998; Logan & Schulkind, 2000). Specifically, $RT_1$ will depend on Task 2 characteristics if $R_1$ is simply held in readiness until $R_2$ is also ready. Thus, the interpretation of a given data set as being consistent or inconsistent with the RSBM can depend critically on whether it is plausible that the data set includes a high enough proportion of trials with grouped responses.

Unfortunately, it is difficult to be sure how much response grouping is actually present in data obtained within a standard PRP paradigm. Grouping can be discouraged to some extent by instructions to respond as quickly as possible to $S_1$ and by the inclusion of long SOAs, but participants might occasionally group their responses anyway. For the most part, researchers attempting to eliminate grouping-based interpretations of their effects have excluded trials with short interresponse times (IRTs). According to the most prominent model of response grouping (Borger, 1963), trials with grouped responses should have especially small IRTs simply because $R_1$ is held until $R_2$ is ready to be emitted too.  

The intuitively plausible procedure of excluding trials with short IRTs is actually somewhat difficult to apply in practice, however, because of two problems. One is that it is not clear exactly how small the IRT should be when responses are grouped, and researchers often have to set somewhat arbitrary cut-offs. For example, Ruthruff, Pashler, and Hazeltine (2003) considered IRTs less than 30 ms to be indicative of grouping, whereas Pashler (1994b) used IRT cut-offs of 75–100 ms, and De Jong (1993) used a cut-off of 200 ms. This is quite a large discrepancy in cut-offs, and it could clearly have a big impact on the proportion of trials that are excluded as having been grouped. Moreover, many of the somewhat arbitrary cut-offs chosen by these and other PRP researchers seem rather large in comparison with IRTs obtained in studies of bimanual responses in other paradigms. For example, Ulrich and Stapf (1984) had participants make bimanual responses in a simple RT task and found that most IRTs were less than 10 ms for one participant and less than 25–35 ms for two others. Similarly, Rinkenauer, Ulrich, and Wing (2001) found most IRTs to be less than 50 ms in their study of bimanual movements, and they concluded that any trials with IRTs greater than that had not been produced simultaneously as instructed. A second problem is that many reasonable models predict that there should occasionally be very small IRTs even if there is actually no response grouping, so exclusion of trials with short IRTs

---

1 Davis (1959) used a different approach, testing for response grouping by examining the correlation of $RT_1$ with $RT_2$ across trials. If participants group responses, then there should be a strong positive correlation; in the limit of simultaneous responses, in fact, the correlation should be perfect. Davis found a correlation of only .77 at an SOA of 50 ms and concluded that this was not high enough to suggest response grouping. More detailed consideration suggests that the $RT_1/RT_2$ correlation is not a very clear-cut diagnostic of response grouping, however (cf. Navon & Miller, 2002; Ulrich & Miller, 2007). For one thing, this correlation need not be extremely strong if responses are only grouped on some small proportion of trials, say 20–50%. Indeed, within some models, grouping responses on a proportion of trials can even decrease the $RT_1/RT_2$ correlation relative to what would be expected with no grouping (Ulrich & Miller, 2007). Moreover, even a small proportion of grouped trials might significantly affect mean RTs, regardless of the resulting $RT_1/RT_2$ correlation. Finally, even a rather high $RT_1/RT_2$ correlation would not necessarily indicate that responses were ever grouped, because there are a variety of other mechanisms that could induce positive $RT_1/RT_2$ correlations (e.g., Navon & Miller, 2002; Pashler, 1994a, 1994b).
could distort the results by excluding some trials without grouping. Under capacity models, for example, the two tasks might be processed in parallel and finish at nearly the same time by chance on some trials. Under central bottleneck models, too, the postbottleneck processes required for Task 1 might occasionally take longer than those for Task 2, again perhaps just due to random variation in processing times, so that the two responses would be emitted at very nearly the same time.

In view of the theoretical importance of response grouping within the debate about PRP models, it seems crucial to investigate both the characteristics of grouped responses and the determinants of response grouping. With respect to the former, attention has so far been focused entirely on the timing characteristics of grouped responses (i.e., short IRTs). Researchers wishing to exclude grouped responses based on short IRTs, for example, would benefit from some empirical guidance in selecting what cut-off IRT value should be used. In addition, though, grouped responses might differ from ungrouped responses with respect to their force–time dynamics as well as their timing, so these dynamics might also be useful in distinguishing between grouped and ungrouped responses. Studies of bimanual motor coordination amply document the fact that it is difficult to perform different movements with the two hands at the same or nearly the same time (e.g., Kelso, Southard, & Goodman, 1979; Marteniuk & MacKenzie, 1980; see Heuer, 1996, for a review). Instead, motor tasks tend to be easiest when the two hands act in concert. For example, people can produce two desired target actions or action sequences more quickly and accurately when the targets share common temporal (e.g., Klapp, 1979; Rinkenauer et al., 2001), spatial (e.g., Franz, Zelaznik, & McCabe, 1991), or force requirements (e.g., Rinkenauer et al., 2001). Based on such phenomena, the two hands are often said to be strongly coupled. It is not yet clear whether this coupling arises because a single mechanism controls both hands (e.g., Franz et al., 1991; Kelso, Putnam, & Goodman, 1983; Kelso et al., 1979; Schmidt, Zelaznik, Hawkins, Frank, & Quinn, 1979) or because two separate hand-specific mechanisms are strongly interlinked (Marteniuk & MacKenzie, 1980; Preilowski, 1975; Spijkers & Heuer, 1995), but the phenomenon of coupling itself is not in doubt. If grouped responses are more strongly coupled than ungrouped responses within PRP paradigms, then, the two responses’ force–time characteristics might be more similar for grouped than ungrouped responses, and such similarity could provide a further criterion for detecting response grouping.

The issue of response grouping is also theoretically relevant beyond the PRP paradigm. Research on this issue may lead to further understanding of the causes of response coupling in bimanual tasks (e.g., Heuer, 1995). Response coupling is usually investigated with simultaneous responses, because it is thought that the tendency to couple the movements of the two hands may originate from a common temporal scheduling mechanism (e.g., Rinkenauer et al., 2001). Such a view might need modification in light of further information about coupling obtained within the PRP paradigm.

Researchers would also benefit from more information about the determinants of response grouping. For one thing, researchers attempting to avoid response grouping in PRP experiments would clearly benefit from information about how the probability of response grouping is influenced by different experimental manipulations. It would perhaps be even better to have a model for the cognitive mechanisms that determine whether participants group their responses in a given trial. Although Borger’s (1963) grouping model describes what happens when responses are grouped (see also Pashler & Johnston, 1989), it does not address the decision of whether such grouping actually takes place in a given trial. Ulrich and Miller (2007) considered a number of response grouping models that included mechanisms for making this decision, including several extensions of Borger’s model. In one model, for example, participants decide probabilistically in advance of each trial whether to group responses...
in that trial. In other models, in contrast, participants decide whether to group based on the SOA, on the relative finishing times of the perceptual analyses of $S_1$ and $S_2$, or on the relative finishing times of Task 1 and Task 2 response selection. They concluded, however, that existing data were not able to discriminate among those models.

Although there have been several previous studies of response grouping in PRP paradigms, none have directly addressed these issues. Instead, these studies have generally been designed to investigate how performance changes when responses are grouped, explicitly instructing participants to group their responses (e.g., Sanders, 1964; Sanders & Keuss, 1969). For example, Pashler and Johnston (1989, Exp. 2) found that the interactions of certain second-task factors and SOA were unchanged when participants were encouraged to group responses, and they interpreted this finding as evidence in favour of a central-bottleneck model. In contrast, De Jong (1993, Exp. 4) found that the interaction of second-task perceptual manipulations with SOA was greatly reduced when participants were instructed to group their responses, which he interpreted as evidence that a response-initiation bottleneck was involved in the production of that interaction in the ungrouped condition. Ruthruff, Pashler, and Klaassen (2001) showed that the difficulty of the easier task still had an effect on RT when responses were grouped, from which they concluded that there must be central as well as peripheral (e.g., response initiation) interference in the PRP paradigm. Sommer, Leuthold, Abdel-Rahman, and Pfütze (1997) used psychophysiological measures with instructed grouping to show that $R_1$ is actually prepared to a motor level and then held in that state until $R_2$ is ready. As these studies illustrate, experiments with instructed response grouping can be used to address various questions about how such grouping affects processing in the PRP paradigm. Yet the fact that participants can group responses when they are instructed to do so reveals little about how participants decide whether to group in PRP paradigms without such instructions. Moreover, it is unclear how well the response characteristics (e.g., IRTs) observed when participants are instructed to group their responses would match up with the characteristics of incidentally grouped responses in standard PRP paradigms without such instructions.

In sum, the present experiments investigated the determinants and characteristics of grouped responses with the general goals of determining what experimental factors influence response grouping, what cognitive mechanisms determine whether grouping takes place, and how grouped responses differ from ungrouped responses. In the first study, we compared bimanual responses in a dual-task paradigm with those obtained in a simple RT task where response grouping would be expected on most if not all trials. SOAs were introduced in Experiment 2, and stimulus order was varied unpredictably in Experiment 3.

**EXPERIMENT 1**

This experiment was a first attempt to investigate the temporal and force–time characteristics of grouped responses. Each participant performed a simple RT task and two independent go/no-go tasks (i.e., a dual-task condition) in different blocks of trials. The basic strategy was to investigate the production of bimanual responses in a dual-task setting, where they might or might not be grouped, as compared with a bimanual simple RT task in which responses should be grouped in virtually every trial. Comparing the IRTs and force–time characteristics of responses observed in the simple RT and dual-task settings should help reveal the characteristics and proportion of grouped responses within the dual-task setting.

In both the simple RT and dual-task blocks, the stimuli were an auditory stimulus, a visual stimulus, both, or neither. When both stimuli were presented, they were always simultaneous. In the simple RT blocks, participants were instructed to make a bimanual response as quickly as possible if any stimulus was presented and to withhold the response (no-go) if no stimulus was presented. We expected that the bimanual responses would be grouped in virtually every trial in this task,
because the two key presses were always made as part of a single unified response that could be prepared in advance and executed as a unit.

In the dual-task blocks, participants were instructed to perform two independent go/no-go tasks—one task with each hand. Specifically, they were instructed to make a key press response with one hand if the auditory stimulus was presented but to withhold that response if no auditory stimulus was presented. They were also instructed to make a key press response with the other hand if the visual stimulus was presented but to withhold that response if no visual stimulus was presented. They were told that sometimes both stimuli would be presented, in which case they should press both keys, and that sometimes neither stimulus would be presented, in which case they should not respond at all. Finally, it was stressed that they should respond as quickly as possible to each stimulus, giving both equal priority.

Because the two responses were made independently in the dual-task blocks, responses could not be fully prepared and executed as a single unit, as they could in the simple RT blocks. Therefore, we expected less response grouping in the dual-task blocks than in the simple RT blocks. Nonetheless, response grouping might occur in some proportion of trials if response grouping is indeed a useful strategy. If responses were sometimes grouped, we expected to find a proportion of grouped trials in the dual-task blocks task that showed characteristics similar to those observed in the simple RT blocks. We measured both response latency and response force in an effort to detect grouped responses.

**Method**

**Participants**

Participants were 12 right-handed first-year students in psychology at the University of Otago who took part in the experiment in partial fulfilment of a course requirement. All had normal or corrected-to-normal vision. Each participant attended a single experimental session lasting about 45 min.

**Apparatus and stimuli**

Stimuli were presented, and responses were recorded by an IBM-PC compatible computer. The visual stimulus was a white square (luminance approximately 70 cd/m²) presented at fixation on the dark background of a standard computer monitor. Viewed from a distance of approximately 60 cm, the square was 2.2° of visual angle on a side, and it remained on the screen until the participant responded. Auditory stimuli were 900-Hz tones of 56 ms duration, and they were presented binaurally over headphones at approximately 50 dB SPL.

Responses were made with the left and right index fingers on two force-sensitive keys similar to telegraph keys. Each key was constructed from a leaf spring (140 × 20 × 2 mm). One end of the spring was fixed in a pedestal, and the other end extended toward the participant at a height of approximately 7 mm above a metal base that provided full forearm support. Participants responded by pressing down the free end using a quick finger flexion, with a cut-out from the metal base allowing essentially unlimited downward movement of the key. Strain gauges (Type 6/120 LY 41, manufactured by Höttinger Baldwin Messtechnik, Darmstadt, Germany) were attached near the fixed end of the leaf spring, so force applied to the free end altered an analogue signal with a resolution of approximately 2.8 mN. A force of 15 N bent the free end of the key by approximately 2 mm. In each trial, the force signal from each key was digitized at 250 Hz, starting 200 ms before the onset of the stimuli and continuing for 2.2 s. Each force key was calibrated using a 50-g weight at the beginning of testing for each participant.

**Procedure**

Each trial began with the display of a plus sign centred at fixation. This warning signal remained on the screen for 800 ms. A period of 200 ms after its offset, a single visual or auditory stimulus was presented, both stimuli were presented, or no stimulus was presented in a catch trial. If no response was made—either because no stimuli were presented or because of a miss—the trial
was terminated 2.2 s after the offset of the warning stimulus.

After the participant responded on a given trial, accuracy feedback was given. If the correct response was made, the word “CORRECT” was displayed for 600 ms. If an error was made, the word “ERROR” was displayed for 1.2 s. The fixation point appeared to begin the next trial approximately 1 s after the offset of this feedback.

Within each task (simple RT versus dual task), each participant was tested first in one short practice block and then in two experimental blocks. The experimental blocks included 16 trials with no stimulus, 24 trials with a visual stimulus, 24 trials with an auditory stimulus, and 36 with both stimuli. Thus, the two stimuli occurred independently, each with a probability of .6. The practice blocks had 1/4 as many trials in each of these stimulus conditions. The order of the trials in each block was randomized separately for each participant. Half of the participants were tested first in the three simple RT blocks and then in the dual-task blocks, whereas the other half of the participants were tested in the reverse order. In the dual-task blocks, half responded to the auditory stimulus with the left hand and to the visual stimulus with the right hand, whereas these assignments were reversed for the other half.

In the simple RT blocks, participants were told to respond by pressing down the response keys with both index fingers if either or both stimuli occurred. In the dual-task blocks, they were told that each response key corresponded to one stimulus, and they should press down the response key corresponding to any stimulus that was presented, responding with both hands if both stimuli appeared. In both the simple RT and dual-task blocks, participants were told to respond as quickly as possible without making false alarms or other errors.

Methods of analysis
The practice blocks were excluded from the analyses. For each trial and hand separately, RT was scored as the latency at which the force generated by that hand reached the criterion level of 100 cN, which is approximately the amount of force needed to depress a key for most standard keyboards, and peak force (PF) was scored as the maximum level of force output produced during the 1,500-ms interval starting at stimulus onset. The IRT for each trial was scored as the absolute value of the difference between $RT_1$ and $RT_2$, and the analogous interresponse difference in peak force value (IPF; i.e., the absolute difference between $PF_1$ and $PF_2$) was also scored. Means of these dependent variables were computed averaging across hands. Correlations of the scores across left and right hands were computed separately for each condition and participant, correlating across trials, pooled across all of the blocks included in the analysis. In all repeated measures analyses of variance (ANOVAs), $p$-values were adjusted using the Greenhouse–Geisser correction for violations of the sphericity assumption (Huynh, 1978) as appropriate.

Results and discussion
In the simple RT blocks, there were 1.0% false alarms in catch trials and 0.2% misses in stimulus-present trials. In the dual-task blocks, there were 0.5% false alarms in catch trials. In addition, participants failed to respond in 0.2% of stimulus-present trials, responded with the wrong hand in 1.3% of single-stimulus trials, incorrectly responded with both hands in 8.1% of single-stimulus trials, and incorrectly responded with only one hand in 1.0% of dual-stimulus trials. Interestingly, when two responses were made to a single stimulus, the IRT was less than 50 ms in 74% of trials, and the first response was the one associated with the nonpresented stimulus in 96% of the trials with IRT >50 ms. Error trials were excluded from subsequent analyses. We also excluded as outliers trials with RTs less than 100 ms (0.3%) or greater than 1,100 ms (1.0%).

Table 1 shows mean RT and PF as a function of the task and stimulus condition. For trials with bimanual responses, it also shows the means of IRT and IPF, as well as the mean correlations of the two hands in RT ($r_{RT}$) and in PF ($r_{PF}$).
Preliminary analyses of RT showed that, as was expected, responses were much faster in the simple RT blocks than in the dual-task blocks, $F(1, 11) = 121.23, MSE = 6,671.3, p < .001$, and there was also a significant interaction of task and stimulus, $F(2, 22) = 15.96, MSE = 665.24, p < .001$. In the simple RT blocks, responses were substantially faster when both stimuli were presented than when just one was, $F(2, 22) = 18.96, MSE = 1,404.6, p < .001$, replicating the redundancy gain typically found in bimodal simple RT tasks (e.g., Diederich & Colonius, 1987).\(^2\) In the dual-task blocks, however, there were no significant differences in RTs to the three different stimuli, $F(2, 22) = 1.56, MSE = 1,192.2, p > .2$. Preliminary analyses of PF revealed no significant effect of task, stimulus, or their interaction (all $p$s > .25).

Of particular interest in connection with response grouping are the differences between tasks in IRT, IPF, $r_{RT}$, and $r_{PF}$, which were evaluated using single-factor ANOVAs comparing the dual-task results from the both-stimulus condition against the simple RT task results averaging across stimulus conditions. The results obtained with these four measures provide strong evidence that responses were less often or less strongly grouped in the dual-task blocks than in the simple RT blocks. IRT was significantly smaller in the simple RT blocks than in the dual-task blocks, $F(1, 11) = 16.14, MSE = 979.54, p < .005$, and so was IPF, $F(1, 11) = 5.71, MSE = 1,096.6, p < .05$. Similarly, there were stronger correlations between the two hands in the simple RT blocks than in the dual-task blocks for both RT, $F(1, 11) = 14.49, MSE = 0.063, p < .005$, and PF, $F(1, 11) = 6.07, MSE = 0.024, p < .05$. Indeed, correlations of $RT_1$ and $RT_2$ were nearly perfect in the simple RT blocks, replicating earlier findings with bimanual responses (e.g., Ulrich & Stapf, 1984).

To investigate in more detail the characteristics of grouped responses, we computed cumulative

---

\(^2\)The race model inequality (Miller, 1982) was also tested to see whether simple RT responses to redundant bimodal stimuli were faster than could be explained in terms of a race model. Using the method described by Ulrich, Miller, and Schröter (in press), tests were carried out for the 5, 15, 25, . . ., 95 percentile points of the cumulative distribution functions (CDFs) for the single- versus redundant-stimulus RT distributions, and the race model was significantly violated ($p < .05$) in the comparisons for all of the percentiles in the range of 15–65.
frequency distributions of the IRT and IPF values for the trials with bimanual responses, simply pooling all such trials across participants. The resulting distributions are shown in Figure 1. The frequency distribution of IRT in the simple RT task is remarkably narrow, with approximately 98% of the IRTs less than 50 ms. This indicates that two preplanned bimanual responses—which should be maximally grouped—can normally be executed within 50 ms of one another, and it suggests that a 50-ms IRT cut-off may be appropriate for identifying grouped responses in the dual-task condition.

The frequency distribution of dual-task IRTs is consistent with the ideas that many—but not all—of the responses are grouped and that IRTs of grouped responses are less than 50 ms in the dual-task blocks, just as they were in the simple RT blocks. Approximately 77% of dual-task IRTs were less than 50 ms, suggesting that responses were grouped in this percentage of the dual-task trials requiring bimanual responses. The IRTs greater than 50 ms were spread out over a wide range, consistent with the idea that these IRTs were generated by ungrouped rather than grouped responses to the two stimuli. In addition, the marked reduction in the slope of the cumulative distribution function (CDF) at 50 ms is consistent with the idea that no grouped responses have IRTs greater than that. Thus, these results suggest that the IRTs of grouped responses were not substantially different in the dual-task blocks, where grouping occurred in only a proportion of trials, than in the simple RT blocks, where it occurred in virtually all trials.

The analogous frequency distributions of IPF for the simple RT and dual-task blocks, also shown in Figure 1, present quite a different picture. Remarkably, the frequency distribution for this measure was not substantially more narrow in simple RT blocks than in dual-task blocks, suggesting that the two hands do not necessarily produce highly similar forces in grouped responses. This result suggests, then, that IPF is a much less useful diagnostic of response grouping than is IRT.3

3 We also analysed response force–time functions in several other ways to look for force measures that would be useful in discriminating between grouped and ungrouped responses, but we had no success at that. For example, several of these analyses compared the full force–time profiles of the two bimanual responses (e.g., by correlating their values across time points), based on the intuition that grouped responses should have more similar profiles than ungrouped responses. In all such analyses, the force-based measures produced only slightly more evidence of grouping in the simple RT blocks than in the dual-task blocks, with largely overlapping distributions like those shown for the IPF measure in Figure 1.
EXPERIMENT 2

The results of Experiment 1 suggest that response grouping can be quite common in dual tasks as well as in bimanual simple RT tasks, but it is possible that the rather high frequency of grouped responses in Experiment 1 may have resulted at least partly from certain special characteristics of that experiment. For one thing, Experiment 1 used exclusively simultaneous presentation of $S_1$ and $S_2$ rather than the sequential presentation used in most PRP studies.

As noted by previous researchers working with the PRP paradigm (e.g., Pashler, 1994a), it seems likely that increasing the SOA would reduce the tendency toward response grouping. Grouping requires the first response to wait until the second is available, and the wait tends to increase with the SOA. Such waiting would be inconsistent with the experimenter's instruction to execute $R_1$ as quickly as possible, especially when SOA is large, so participants should avoid grouping with large SOAs. Thus, it is quite possible that a simple modification of Experiment 1 to include SOAs much greater than zero would reduce grouping substantially or eliminate it altogether.

In this experiment, fixed-SOA blocks with SOA = 0 were compared with varied-SOA blocks in which stimuli were presented with SOAs of 0, 100, 200, 300, or 400 ms, again using the dual go/no-go tasks of Experiment 1. Based on the idea that response grouping should be less common with longer SOAs, it was expected that there would be less response grouping in the varied blocks than in the fixed blocks.

If response grouping is reduced in varied blocks as expected, the reduction might take either of two forms, and the two possible outcomes would have somewhat different theoretical interpretations that might be termed “advance-decision grouping” versus “online-decision grouping”. One possible outcome is that response grouping would be reduced at all SOAs in the varied blocks, even at SOA = 0, and this outcome would support advance-decision grouping. If participants decide whether to group in advance of the trial, as is assumed by one reasonable extension of Borger's (1963) model (cf. Ulrich & Miller, 2007), then the probability of grouping must be independent of SOA when SOAs are randomized within blocks. A second possibility is that the probability of response grouping would decrease as SOA increases, and this outcome would indicate online-decision grouping. That is, given that the SOA was not known in advance, such a dependence would indicate that the decision of whether to group responses was made during the trial itself. A number of different grouping models are consistent with online-decision grouping, including some extensions of Borger’s model (cf. Ulrich & Miller, 2007).

Method

Participants were 16 new volunteers recruited from the same population as that tested in Experiment 1. Each was tested in two practice blocks and four experimental blocks during a single experimental session lasting approximately 45 min. For one experimental block, the procedure was exactly the same as that in the dual-task blocks of Experiment 1. For the other three experimental blocks, the redundant-stimulus trials were equally divided among five different SOA conditions: 0, 100, 200, 300, or 400 ms. Each experimental block included 104 trials—16 catch, 24 visual, 24 auditory, and 40 redundant—and there were one quarter as many trials in each practice block. For each participant, the same stimulus (i.e., auditory or visual) was always presented first in the redundant-stimulus trials with positive SOAs, so stimulus order was predictable. Across four groups of participants, we counterbalanced whether the tone or the light was presented first on trials with positive SOAs and also counterbalanced whether the left hand was used to respond to the tone and the right hand to the light, or vice versa. It is convenient to use $RT_1$ to refer to the response time to whichever stimulus appeared first on trials with positive SOAs, to use $RT_2$ to refer to the response time to the other stimulus, and to distinguish analogously between PF1 and PF2.
Results and discussion

Participants incorrectly responded in 1.0% of catch trials, and they incorrectly failed to respond at all in 0.4% of stimulus-present trials. They responded with the wrong hand in 0.7% of single-stimulus trials, incorrectly responded with both hands in 6.9% of single-stimulus trials, and incorrectly responded with only one hand in 0.6% of dual-stimulus trials. Error trials were excluded from subsequent analyses, as were trials with RTs less than 100 ms (0.0%) or greater than 1,100 ms (1.5%).

Mean reaction time and peak force

The upper panel of Figure 2 displays the mean RTs as a function of experimental condition. Preliminary ANOVAs on these RTs compared the fixed versus varied blocks with respect to the two SOAs included in both blocks (i.e., SOA = ∞ versus SOA = 0). These ANOVAs yielded no significant main effects or interactions involving block type or SOA for either RT1 or RT2 (all ps > .1).

The main ANOVAs focused on the effect of SOA within the varied blocks, excluding the single-stimulus condition (i.e., SOA = ∞). As is virtually always found in PRP paradigms, RT2 decreased as SOA increased, $F(4, 56) = 28.97, MSE = 1,119.9, p < .001$. This strong dependence of RT2 on SOA is one of the classic signatures of the PRP task, so its presence confirms that the task combination used here is comparable to many other task combinations that have been studied previously.

Furthermore, RT1 increased significantly as SOA increased, $F(4, 56) = 15.53, MSE = 2,201.9, p < .001$, consistent with the idea that response grouping slows RT1 especially at longer SOAs. As noted in the Introduction, an increase in RT1 with SOA is one of the phenomena inconsistent with the RSBM that is often attributed to grouping.

The lower panel of Figure 2 displays the mean PFs. ANOVAs were carried out on PF1 and PF2 parallel to those carried out with RT1 and RT2. In the preliminary ANOVAs comparing fixed versus varied blocks, the only significant effect was that responses were more forceful to single stimuli (i.e., SOA = ∞) than to simultaneous double stimuli (i.e., SOA = 0) for PF1, $F(1, 14) = 7.88, MSE = 1,1587.4, p < .025$. It thus seems that the appearance of the second stimulus draws off some sort of energy or capacity that would otherwise contribute to the forcefulness of R1. The analyses of SOA effects in the varied blocks indicated only a significant effect of SOA on PF2, $F(4, 56) = 3.27, MSE = 756.03, p < .05$, but we suspect this may have been a Type I error because the mean PF2 values did...
not vary monotonically and because this effect was not replicated in the next experiment.

*Measures of grouping*

Figure 3 shows mean IRT, IPF, $r_{RT}$, and $r_{PF}$ for the bimanual response trials as a function of fixed versus varied block and of SOA. Differences between fixed and varied blocks with SOA = 0 were rather small, approaching significance only for IPF, $F(1, 14) = 3.26$, $MSE = 3,217.9$, $p < .1$. Within the varied blocks, SOA had highly significant effects on both IRT, $F(4, 56) = 41.22$, $MSE = 2,182.5$, $p < .001$, and $r_{RT}$, $F(4, 56) = 7.37$, $MSE = 0.074$, $p < .001$, strongly suggesting that response grouping decreased with increasing SOA, as expected. Interestingly, $r_{RT}$ does not decrease monotonically with SOA, which is consistent with the finding that $r_{RT}$ is not monotonically related to the proportion of grouped trials within some grouping models (Ulrich & Miller, 2007). SOA had no significant effect on either IPF or $r_{PF}$ (both $p s > .17$), reinforcing Experiment 1’s finding that PF measures are relatively insensitive to response grouping.

Figure 4 shows the CDFs of IRT and IPF values as a function of block and SOA. As in

*Figure 3. Mean interresponse time (IRT, in ms), between-hand difference in peak force (IPF, in cN), and correlations between Task 1 and Task 2 in reaction times ($r_{CT}$) and peak forces ($r_{PF}$), as a function of fixed versus varied block and stimulus onset asynchrony (SOA) in Experiment 2.*
Experiment 1, the CDFs of IRT in all conditions show a rapid rise in the region from 0–50 ms, suggesting a substantial proportion of trials with grouped responses with IRTs in that range, followed by a slower rise presumably reflecting IRTs in ungrouped trials. Once again, no comparable inflections are present in the CDFs of IPF, suggesting that this measure is not useful in identifying grouped responses.

Figure 5 shows the proportion of trials with IRT ≤ 50 ms as a function of block and SOA, which seems—to be an appropriate index of grouping probability. The simplest interpretation of this pattern is that participants produce a mixture of grouped and ungrouped responses, with grouping probability decreasing as SOA increases and, at least for SOA = 0, decreasing when SOA varies randomly within a block rather than being fixed. As was discussed in the introduction to this experiment, this pattern strongly supports online-decision grouping over advance-decision grouping.

Finally, Figure 6 shows values of mean RT₁ and RT₂ on dual-stimulus trials, computed separately for trials with short versus long IRTs (i.e., IRT ≤ 50 ms versus IRT > 50 ms).4 As noted earlier, exclusion of trials with short IRTs is an analysis strategy that is often used in an effort to remove the effects of response grouping (e.g., De Jong, 1993; Hommel, 1998; Lien, Schweickert, &
It is interesting to note that the overall mean $RT_1$ shown in Figure 2 increased with SOA, despite the fact that grouping probability decreased with SOA (Figure 5). For reasons discussed earlier, grouped responses produce especially long $RT_1$ values at larger SOAs (cf. Figure 6, lower panel). Clearly, at longer SOAs the increase in grouped $RT_1$ values is more than enough to make up for the decreased proportion of grouped trials, resulting in the overall increased mean shown in Figure 2.

Conclusions
As in Experiment 1, grouped responses appeared to be identifiable in terms of their IRTs (i.e., as having IRT $\leq 50$ ms) but not in terms of similarity of their force characteristics. By this criterion, responses were often grouped—more than 75% of the time with a fixed SOA = 0. The frequency of response grouping clearly decreased as SOA increased, but grouping nonetheless caused mean $RT_1$ to increase substantially at longer SOAs.

The present results provide some clues about the mechanisms determining response grouping in this experiment. First, the frequency distributions of IRT indicate clearly that response grouping is a probabilistic phenomenon, with responses being grouped in some trials but not others. Second, the dependence of grouping probability on SOA (e.g., Figure 5) indicates that response grouping is determined online. That is, participants do not simply decide in advance of the trial whether to group, but instead decide whether to group during the processing that takes place after the trial begins.

EXPERIMENT 3

Another experimental manipulation that might well influence the likelihood of response grouping, in addition to SOA, is to vary the order of stimulus presentation. Pashler (1990) found that unpredictable response order greatly increased the difficulty of PRP tasks with two manual responses, and he concluded that such unpredictability increased the sources of interference in specifying two
manual responses. Participants in tasks with unpredictable response order might reasonably be expected to minimize grouping to avoid this interference. In this experiment, then, participants were tested in a dual-task condition in which dual stimuli could be presented in either order as well as simultaneously. Denoting the conditions with the auditory stimulus first by positive SOAs and those with the visual stimulus first by negative SOAs, the SOAs used in dual-stimulus trials in this experiment were –400, –200, 0, 200, and 400 ms.

**Method**

Participants were 16 new volunteers recruited from the same population as that tested in Experiments 1 and 2. The procedure was nearly identical to that of Experiment 2 except for the SOAs, which were equally divided among the values of –400, –200, 0, 200, and 400 ms, measured from the onset of the auditory stimulus to the onset of the visual stimulus. Note that stimulus order was unpredictable because of the mixing of positive and negative SOAs. The assignments of visual and auditory stimuli to the left and right hands were counterbalanced across participants, as was the order of testing in the blocks with fixed versus varied SOAs.

**Results**

Participants incorrectly responded in 0.1% of catch trials, and they incorrectly failed to respond at all in 0.1% of stimulus-present trials. They responded with the wrong hand in 0.9% of single-stimulus trials, incorrectly responded with both hands in 4.5% of single-stimulus trials, and incorrectly responded with only one hand in 0.7% of dual-stimulus trials. Error trials were excluded from subsequent analyses, as were trials with RTs less than 100 ms (0.0%) or greater than 1,100 ms (1.6%).

**Figure 7.** Mean reaction time (RT, in ms, upper panel) and peak force (PF, in cN, lower panel) in Tasks 1 and 2 as a function of fixed versus varied block and stimulus onset asynchrony (SOA) in Experiment 3. SOAs of ∞ denote trials in which only one stimulus was presented.

**Mean reaction time and peak force**

The upper panel of Figure 7 displays the mean RTs as a function of experimental condition. Preliminary ANOVAs again indicated no difference between fixed and varied blocks for RTs in the conditions common to both blocks (i.e., SOA = ∞ versus SOA = 0; ps > .1). Within

---

5 Because the SOAs were symmetric around zero, the distinction between Tasks 1 and 2 is arbitrary for single-stimulus trials and for trials with SOA = 0. Therefore, for these conditions identical Task 1 and Task 2 mean RTs were computed averaging across both the visual and auditory stimuli.
the varied-SOA blocks, the main effect of SOA was highly significant for both RT\(_1\), \(F(2, 28) = 8.52, MSE = 3,727.8, p < .01\), and RT\(_2\), \(F(2, 28) = 38.72, MSE = 909.50, p < .001\), replicating the trends observed in Experiment 2.

The lower panel of Figure 7 displays the mean PFs. As is evident from the figure, force was larger in fixed blocks than in varied blocks, \(F(1, 14) = 10.80, MSE = 18,404.6, p < .01\), suggesting that even the potential later appearance of a second stimulus draws off some sort of energy or capacity that would otherwise contribute to the forcefulness of R\(_1\). Experiment 2’s finding of greater forcefulness to single than dual stimuli was again present in the means, but it only approached statistical reliability, \(F(1, 14) = 3.00, MSE = 4,226.1, .10 < p < .11\). The analyses of SOA effects in the varied blocks indicated no significant effect of SOA on either PF\(_1\) or PF\(_2\), \(p > .15\).

**Measures of grouping**

Figure 8 shows mean IRT, IPF, \(r_{RT}\), and \(r_{PF}\) for the bimanual response trials as a function of fixed versus varied block and of SOA. Differences between fixed and varied blocks with SOA = 0 were not reliable for either of the
correlation-based measures, but they were reliable for both IRT, \( F(1, 14) = 11.92, \text{MSE} = 267.93, p < .01 \), and IPF, \( F(1, 14) = 4.83, \text{MSE} = 667.55, p < .05 \). Mean IRT was smaller in fixed blocks than in varied blocks, consistent with the idea that there should be more response grouping in the former. In contrast, mean IPF was larger in fixed blocks than in varied blocks, giving yet another sign that force measures are not particularly sensitive to response grouping. Within the varied blocks, SOA had highly significant effects on both IRT, \( F(2, 28) = 43.27, \text{MSE} = 3,369.0, p < .001 \), and \( r_{RT} \), \( F(2, 28) = 40.35, \text{MSE} = 0.0459, p < .001 \), again strongly suggesting that response grouping decreased with increasing SOA. Unlike Experiment 2, SOA had significant effects on both IPF, \( F(2, 28) = 5.53, \text{MSE} = 362.44, p < .025 \), and \( r_{PF} \), \( F(2, 28) = 6.54, \text{MSE} = 0.0128, p < .01 \), but these effects were numerically quite small in comparison with the corresponding effects on IRT and \( r_{RT} \).

Figure 9 shows the CDFs of IRT and IPF values as a function of block and SOA. As in both of the earlier experiments, the CDFs of IRT rise rapidly from 0–50 ms, suggesting the presence of grouped responses, followed by a slower rise presumably reflecting in ungrouped trials. Once again, the CDFs of IPF show no comparable pattern suggesting any special IPF values associated with grouped responses.

Figure 10 shows the proportion of trials with IRT ≤ 50 ms as a function of block and SOA. This proportion declines rapidly as a function of SOA as in Experiment 2, again supporting online-decision grouping.

Finally, Figure 11 shows values of mean \( RT_1 \) and \( RT_2 \) computed separately for trials with short versus long IRTs, and the results nicely replicate those of Experiment 2. Again, the very strong effect of SOA on \( RT_1 \) in trials with short IRTs is consistent with the predicted effects of grouping. Furthermore, the substantially smaller effect of SOA on \( RT_1 \) in the long-IRT trials of

---

6 A total of 4 participants were excluded from the computation of means with long IRTs, and 4 others were excluded from the computation of means with short IRTs, because they had no trials with appropriate IRTs in one or more conditions.
GENERAL DISCUSSION

The results of these three experiments provide a reasonably clear picture of response grouping in an easy dual task with manual responses. We first summarize the observed characteristics of grouped responses (i.e., how they differ from ungrouped responses) and then turn to results concerning the experimental manipulations that affect grouping, models for response grouping, and the methodological implications. A final section discusses the limitations of these results and the need for further investigations.

Characteristics of grouped responses

Consistent with the common assumption of PRP researchers, it seems clear that grouped responses are distinctive by virtue of having small IRTs. The results of these experiments suggest that grouped bimanual responses are normally made within 50 ms of one another, whereas ungrouped responses can be separated by hundreds of milliseconds. Although the exact cut-off for grouped responses may vary slightly depending on the nature of the response sets, it seems unlikely from these results that grouped manual key press responses would ever be separated by more than 100 ms, although grouping cut-offs this large or larger have been used by some researchers.

It is disappointing and somewhat surprising that the similarity of force profiles did not vary much between grouped and ungrouped responses. Given the strong evidence of force coupling with bimanual responses (e.g., Rinkenauer et al., 2001), it seemed plausible to suppose that we could better identify grouped responses by looking for highly similar force–time profiles in addition to small IRTs. Unfortunately, we were
not able to find any profile similarity measures that provided much support for this conjecture, suggesting that force–time profile similarity is not a useful additional indicator of grouping. Informal analysis suggested that part of the reason is that force–time profiles are highly stereotyped across trials even without grouping. For example, the median correlations of randomly paired force–time profiles from different trials were .974–.986 for 11 of the 12 participants in Experiment 1, once these profiles were adjusted to have equal onset times. To the extent that participants produce responses with almost the same force–time profile in every trial anyway, one would not expect minor fluctuations in these profiles to be well correlated between simultaneous bimanual responses.

Experimental manipulations affecting grouping

The results of Experiments 2 and 3 clearly indicate that response grouping is quite sensitive to SOA. Specifically, grouping can be very common at short SOAs, at least with typical instructions emphasizing rapid responses in both tasks, but it decreases fairly rapidly as SOA increases. The dependence of grouping probability on SOA supports online-decision grouping models, in which the decision to group responses is not made in advance of the trial but instead depends on the temporal sequence of events in the trial.

Interestingly, although the probability of response grouping depends strongly on the actual SOA, this probability seems to depend rather little on the set of potential SOAs (cf. Miller, Ulrich, & Rolke, in press), as indicated by the negligible differences between fixed-versus varied-SOA blocks in Experiments 2 and 3, even when stimulus order was uncertain (i.e., Experiment 3). This suggests that advance preparation for particular expected SOAs is not very important in determining grouping, with grouping instead determined mainly by stimulus- and processing-driven events taking place during the trial itself.

Models for response grouping

Although these experiments were not explicitly designed to test among different formal models of response grouping (Ulrich & Miller, 2007), several aspects of the results do have implications for such models.

As a baseline model, we first consider the possibility that there is actually no response grouping at all. According to this model, short IRTs arise naturally from the separate processes handling the two tasks, because it sometimes happens that the two processes complete at about the same time. Furthermore, assuming that the two processes have approximately the same duration, short IRTs would naturally be less frequent as SOA increased, just because this would decrease the probability of fortuitous near-simultaneous completion. The main problem with this baseline model is that it provides no explanation for the observed increase in $RT_1$ with increasing SOA. Clearly, this observed pattern suggests that $R_1$ is being held back for grouping, contrary to the baseline model’s premise of separate responses.

Given the clear existence of some response grouping, what can be concluded about the nature of this grouping? First, as noted earlier, the decrease in grouping with increasing SOA indicates that the decision of whether to group responses is made online, depending at least on the time between stimuli and possibly also on the sensorimotor processing time for each stimulus (e.g., its perceptual detection latency). Second, the results provide clear suggestions of benefits associated with response grouping. The likelihood of grouping benefits has often been cited as a possible reason for grouping (e.g., Ulrich & Miller, 2007), but to our knowledge no such benefits have ever actually been demonstrated. In the current results, such benefits are perhaps most strongly suggested by the comparisons of mean RTs for grouped (i.e., $IRT \leq 50$ ms) versus ungrouped (i.e., $IRT > 50$ ms) responses that can be seen in Figures 6 and 11. With $SOA = 0$, for example, grouped responses are approximately 50 ms faster, on average, than ungrouped responses. This strongly suggests that grouping
allows responses to be emitted faster, although it is possible that the causality works in the opposite direction. That is, it is possible that faster responses are easier to group, so that responses are grouped because they are faster, rather than being faster because they are grouped.

Methodological implications

The major effects of grouping are on $RT_1$, as would be expected based on the intuition that $R_1$ must wait for $R_2$ to be ready on grouped trials. Indeed, response grouping has been suggested as a post hoc explanation of an observed increase in $RT_1$ with SOA, contrary to bottleneck models' prediction that $RT_1$ should be independent of SOA, and the present results support this suggestion. Although the current results indicate that the probability of grouping decreases rapidly with increasing SOA (cf. Figures 5 and 10), the mean $RT_1$ increases with SOA more than fast enough to offset this decreasing probability. Of course it is possible that under other circumstances the grouping probability might decrease much faster, causing $RT_1$ to be independent of SOA or even to decrease with increasing SOA. Tentatively, though, we conclude that response grouping is probably responsible for most observed increases in $RT_1$ with increasing SOA. Fortunately, the effects of grouping on $RT_1$ can be largely eliminated if the data are reanalysed excluding grouped responses (i.e., trials with IRT $\leq$ 50 ms; Figures 6 and 11).

Limitations and future directions

Although the present results give a reasonably clear picture of response grouping with the go/no-go tasks studied here, it is an open question whether a similar picture would emerge with other task pairings. The probability of response grouping surely depends on task characteristics in addition to those investigated here. For example, grouping probability should decrease as Task 2 difficulty increases, for much the same reason as it decreases as SOA increases. Both of these manipulations cause extra increases in $RT_1$ by requiring $R_1$ to be held back longer for grouping with $R_2$. In addition, it seems very plausible that the characteristics of grouped responses might depend on the response set. For example, grouped responses might have substantially larger IRTs when the two responses are less similar in character (e.g., hand and foot, manual and vocal), because more similar responses are likely to be more strongly coupled than those in different modalities. It is also possible that grouped responses might be sensitive to the set of response alternatives, holding constant the physical responses themselves. In the present experiments, for example, a go/no-go task was performed with each hand, so the two potentially grouped responses were always made with the same two fingers (i.e., left and right index). The characteristics of grouped responses might be somewhat different (e.g., larger IRTs) if two-alternative forced-choice tasks were performed by the two hands, so that in different trials different pairs of response fingers might have to be grouped (e.g., left index and right middle). In short, there are still many open empirical questions about how the characteristics and determinants of grouped responses depend on the dual-task situation under study. It seems clear, however, that answers to such questions would greatly increase our understanding of the results obtained from such studies.

REFERENCES


Copyright of Quarterly Journal of Experimental Psychology is the property of Psychology Press (UK) and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.