Running head: Dual Memory Retrieval Practice

To appear in Memory & Cognition

The Specificity of Learned Parallelism in Dual Memory Retrieval

Tilo Strobach[1,2*], Torsten Schubert[1], Harold Pashler[3], & Timothy Rickard[3]

[1] Humboldt-University, Berlin, Germany

[2] Ludwig-Maximilians-University, Munich, Germany

[3] University of California, San Diego, USA

*Corresponding author

Tilo Strobach

Humboldt University Berlin

Department of Psychology

Rudower Chaussee 18

12489 Berlin

Germany

email: tilo.strobach@hu-berlin.de

Phone: +49 (0) 30 2093 4911

Fax: +49 (0) 30 2093 4912

Abstract

Retrieval of two responses from one visually-presented cue occurs sequentially at the outset of dual-retrieval practice. Exclusively for subjects who adopt a mode of grouping (i.e., synchronizing) their response execution, however, response times after dual-retrieval practice indicate a shift to learned retrieval parallelism (e.g., Nino & Rickard, 2003). The present study investigates how this learned parallelism is achieved and why it appears to occur only for subjects who group their responses. Two main accounts are considered: a task-level account vs. a cue-level account. The *task-level account* assumes that learned retrieval parallelism occurs at the level of the task as a whole and is not limited to practiced cues. Grouping response execution may thus promote a general shift to parallel retrieval following practice. The *cue-level account* states that learned retrieval parallelism is specific to practiced cues. This type of parallelism may result from cue-specific response chunking that occurs uniquely as a consequence of grouped response execution. The results of two experiments favor the second account and are best interpreted in terms of a structural bottleneck model.

Introduction

Improved performance following dual-task practice has been observed for a variety of tasks, including simple choice reaction-time (RT) tasks (e.g., Hazeltine, Teague, & Ivry, 2002; Ruthruff, Van Selst, Johnston, & Remington, 2006; Strobach, Liepelt, Schubert, & Kiesel, 2012a), two continuous tasks (e.g., Hirst, Spelke, Reaves, Caharack, & Neisser, 1980), two working memory updating tasks (e.g., Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Oberauer & Kliegl, 2004), and cued recall tasks (e.g., Nino & Rickard, 2003). Identification of the mechanisms that give rise to this improvement is important from the perspectives of dual-task performance models (Logan & Gordon, 2001; Pashler, 1994), models of the human cognitive architecture and its sequential vs. parallel processing characteristics (Meyer & Kieras, 1997a; Townsend & Wenger, 2004), and theories of learning under dual-task conditions (Kramer, Larish, & Strayer, 1995; Nino & Rickard, 1993).

Liepelt, Strobach, Frensch, and Schubert (2011; Strobach, Frensch, Soutschek, & Schubert, 2012b) have addressed this issue of improved dual-task performance for the case of two concurrently practiced choice RT tasks. In their study, subjects practiced two tasks in (1) either dual-task and single-task situations or (2) exclusively in single-task situations. The combined single-task and dual-task practice resulted in a *dual-task learning effect*, as indicated by more improvement in performance on the dual task than could be accounted for by improvement during single-task practice. Importantly, this learning effect was not specific to the choice tasks that were performed during the dual-task practice phase. Rather, it generalized to new stimuli and stimulus-response mapping rules. Liepelt et al. argued that dual-task learning in choice RT tasks takes the form of a generalizable improvement in task coordination, and more specifically, a decreased task switching delay in the context of continued sequential access to a central task processing bottleneck (Pashler, 1994; Schubert, 1999). Studies of extensive practice with simultaneous choice RT tasks favor the possibility of a central bottleneck that prevents parallel stimulus-response execution, is structural and

stubborn (Ruthruff et al., 2006; Van Selst, Ruthruff, & Johnston, 1999) and perhaps even immutable (Anderson, Taatgen, & Byrne, 2005; Ruthruff, Johnston, Van Selst, Whitsell, & Remington, 2003; Schubert, 2008).

In this paper we address the issue of dual-task learning and its characteristics for the case of cued recall. We build on an investigation by Nino and Rickard (2003, Experiment 2; see also Rickard & Pashler, 2005) of two memory retrievals from a single cue. Nino and Rickard's subjects first learned to retrieve a keypress and a vocal-digit response for each of a set of ten color-word cues. In the keypress learning phase, subjects learned to press a left or a right key for each cue. In the vocal-digit learning phase, they learned to speak a unique digit for each cue. For example, upon seeing the word "*red*", a given subject might have learned to press the left response key (i.e., single-retrieval keypress blocks) and, on separate trials, to say the word "*five*" (i.e., single-retrieval vocal-digit blocks). In a subsequent dual-retrieval test phase, subjects were presented with a series of 30 triads, each including one single-retrieval keypress block, one single-retrieval vocal-digit block, and one dual-retrieval block (i.e., each triad includes three different block types). In the dual-retrieval blocks, subjects executed both the keypress and vocal-digit responses on each trial and cue presentation.

The dual-retrieval results of that experiment were best understood through separate analyses of two sets of subjects. One set of subjects, termed response *non-grouper subjects*, had a large mean inter-response interval (IRI) on dual-retrieval trials, and thus appears to have adopted a mode of executing each response sequentially as soon as it was retrieved. In particular, both RT1 (latency between cue presentation and the first executed response) and RT2 (latency between cue presentation and the second executed response) were initially (i.e., Triad 1) longer than could be predicted by a sequential retrieval model that assumes (1) a bottleneck exclusively at a memory retrieval stage of processing, and (2) maximal efficiency in retrieval scheduling and coordination. That model, described in Appendix A, will henceforth be referred to as the *efficient sequential (ES) retrieval model*. The ES model

constitutes a lower bound RT prediction for sequential retrieval; if that lower bound is violated, then the sequential model can be eliminated. Starting at about Triad 5, however, both RT1 and RT2 converged on and closely tracked the quantitative predictions of the ES retrieval model. That pattern suggests that for non-grouper subjects (1) the memory retrieval operations continued to operate sequentially throughout dual-retrieval practice, but (2) some form of coordinative dual-retrieval learning occurred, allowing performance to converge on ES prediction with practice.

Intriguingly, there was a second set of subjects, termed response *grouper subjects*, who had a small mean IRI on dual-retrieval trials. By the beginning of dual-retrieval practice, those subjects appear to have adopted a mode of waiting until they had retrieved both responses before synchronously executing them. As was the case for non-grouper subjects, RT2 for grouper subjects was thus initially longer than the ES retrieval prediction. By the end of practice, however, RT2 fell below its ES prediction by several hundred milliseconds (i.e., the lower bound RT prediction for ES retrieval was violated and the associated sequential model could be eliminated), approaching the predictions of a parallel model that assumes independent and capacity unconstrained retrieval (e.g., a *race model*; see derivation of predictions in Appendix A). Thus, for grouper subjects, some type of *learned retrieval parallelism* occurred.

The results for grouper subjects raise the possibility that the persistent central processing bottleneck, which as noted above appears to hold sway even after extensive practice for choice RT dual tasks (e.g., Ruthruff et al., 2003), does not apply to the case of two memory retrievals from a single cue. Nino and Rickard (2003), however, suggested an alternative account, according to which *cue-level response chunking* occurs during practice, but only when subjects group (i.e., synchronize) their response execution. Because response grouping presumably yields concurrent activation of both task sets (i.e., retrieve the keypress response and retrieval the vocal-digit response) and both types of response information (i.e.,

the keypress and vocal responses) in working memory, response chunking is plausible. In contrast, for non-grouper subjects, several hundred milliseconds intervened between the first and the second response execution. For those subjects, the activation of the first executed response may have fallen to levels too low for chunking to occur.

One strong prediction of Nino and Rickard's (2003) model is that learned retrieval parallelism (demonstrated by dual-retrieval RTs violating the ES lower bound) is specific to the practiced cue-response combinations. This specificity assumption is a consequence of the chunking characteristics: Chunking is exclusively realized for cue-response combinations that are practiced under dual-retrieval conditions. That prediction is tested for the first time here.

In the present study, as in Nino and Rickard (2003), subjects were trained to criterion on a set of cues (i.e., cue-response associations) and then given interleaved single and dual-retrieval practice. Unlike the Nino and Rickard study, however, only half of the set's cues were presented during that practice phase (i.e., old cues) while the remaining cues were not presented. In this way, there is no covert dual-retrieval available for these latter cues while there is overt dual-retrieval on old cues; this is essential to separating overt vs. covert effects (Smith, Roediger, & Karpicke, in press) in dual-retention and dual-retrieval practice. On a subsequent transfer test, the remaining cues of this set (i.e., cues not presented during that single and dual-retrieval practice phase, i.e., new dual-retrieval cues) were included in the absence of the old cues (Experiment 1) or were mixed with the old cues (Experiment 2). The cue-specific response chunking model unambiguously predicts that, for grouper subjects, learned retrieval parallelism will not transfer to the new cues in either experiment; that is, it predicts that the dual-retrieval RTs for new cues on the transfer test will not violate (i.e., fall below) the ES prediction on at least a first transfer block (on which no prior dual response chunking could have occurred for new cues).

Alternatively, if learned parallelism *does* transfer to new cues, then the cue-specific chunking model can be rejected, at least as the primary mechanism of learned retrieval

parallelism. Rather, a *task-level mechanism* of retrieval parallelism would be implied. That task-level mechanism could be interpreted in the context of recent proposals that processing bottlenecks are under participants' strategic control (Meyer & Kieras, 1997b; Oberauer & Bialkova, 2011; Oberauer & Kliegl, 2004; Ruthruff, Pashler, & Klaassen, 2001) or might reflect individual, trait-like peculiarities (Watson & Strayer, 2010); in both cases, subjects choose whether to do sequential or parallel retrieval depending on factors such as their overall confidence with the task, which would be expected to increase with practice. In the task-level account, the response grouping observed for grouper subjects need not be interpreted as promoting cue-level response chunking. Rather, it could promote a shift (either due to strategic control or trait) from sequential to parallel retrieval soon after the beginning of dual-retrieval practice, giving rise to dual-retrieval RTs below the ES prediction. This task-level mechanism should, by definition, yield generalization of learned parallelism to new dual-retrieval cues on the transfer test and hence dual-retrieval RTs that violate the ES lower bound prediction, even on a first transfer block of Experiment 1 and 2.

Note that the case of two retrievals from one cue, as opposed to the more typical design in dual-task research involving two retrievals from two cues, eliminates discrimination of and selection between two presented cues as a source for a processing bottleneck. Further, the use of one instead of two cues reduces or eliminates potential impacts of divided attention within and across modalities (Fargot & Pashler, 1992). Thus, the case of a single cue and two responses has important advantages with respect to task sensitivity to the underlying retrieval stage dynamics that are of interest here (Tombu & Jolicoeur, 2004; Meyer & Kieras, 1997b). In fact, this situation potentially enables the identification of latent processing characteristics of practiced dual-retrieval and its sequential vs. parallel performance (Meyer & Kieras, 1997a; Townsend & Wenger, 2004).

Experiment 1

In Experiment 1 we investigated the cue-specificity of learned parallelism and explored whether this parallelism for grouper subjects transfers to new dual-retrieval cues following practice on old cues, controlling for an equal amount of prior single-retrieval experience on both types of cues. If learned retrieval parallelism reflects a task-level shift from sequential to parallel retrieval (i.e., task-level account), then that learning should transfer to an identical task context with new dual-retrieval cues in a transfer phase, leading to a violation of the ES model's RT2 predictions. If, however, learned parallelism during the practice phase is cue-specific (i.e., if the cue-level chunking account is correct), then dual-retrieval performance on new cues should not violate the ES model (i.e., cue-level account).

Methods

Subjects

Twenty-four undergraduates at the University of California, San Diego, participated in the experiment for partial fulfillment of a course requirement.

Apparatus and Stimuli

Subjects were tested on IBM-compatible personal computers and experiments were controlled by the experimental software package E-Prime software (Psychology Software Tools, Pittsburgh, PA). Vocal-digit responses and manual keypress responses were recorded with the accompanying voice-key apparatus (Model 200A). The list of cues, the corresponding keypress responses in the keypress task, and the vocal-digit responses in the vocal task, are shown in Table 1. Stimulus words subtended up to 7 cm and letter height was 1.7 cm. Stimuli were presented on a 19" CRT monitor.

Insert Table 1 here.

Procedure and Design

An overview of the design is given in Table 2. In the *study phase* for the vocal task, subjects were instructed to memorize the 14 cues and the associated digits. Each of the 14 cue-response combinations was presented once, randomly ordered, in each of two study blocks. On each trial, the cue and the digit were presented for 5,000 ms in the centre of the screen, followed by a blank interval of 1,000 ms and then the presentation of a fixation cross for 500 ms. Next, the cue just previously shown was presented without the digit and subjects were instructed to speak the associated digit clearly into the microphone.

In each block of the subsequent *single-retrieval criterion phase* for the vocal task, each cue was again presented once and subjects were instructed to retrieve the earlier associated response. On each trial, a blank screen appeared for 1,000 ms, followed by a fixation cross for 500 ms and then the presentation of a cue in the centre of the screen until the subjects responded. A blank screen of 2,500 ms then appeared, during which the experimenter entered the subject's vocal-digit response via the keyboard number pad in single-vocal blocks. If the response was correct, the next trial began immediately thereafter. If the response was incorrect, an "incorrect" message, plus the correct answer, was presented for 2,500 ms. At the end of each block of this phase, proportion correct and the mean RT on correct trials was presented. These criterion phase blocks were continued until subjects completed two successive blocks with 100 % accuracy and a mean RT of 1,200 ms or below. Next, identical study and criterion phases were conducted for the keypress task. Half of the subjects completed these two phases for the vocal task first, and half for the keypress task first.

In the *single-retrieval practice phase*, ten blocks of the vocal and the keypress tasks were presented in alternating order, starting with a block of the vocal task for half of the subjects and a block of the keypress task in the remaining subjects. On each single-retrieval practice block, all 14 cues were presented once. Trials were identical to trials in the single-

retrieval criterion phase. This phase assured a high level of accuracy and short RTs for the single tasks going into the single-dual practice phase.

The *single-dual practice phase* consisted of 20 triads (Practice triads 1 – 20), each consisting of three blocks of seven trials, one trial each for seven of the previously trained 14 cues. In each triad, half of the subjects performed the vocal task in the first single-retrieval block, the keypress task in the second single-retrieval block, and dual retrievals in the third block in each triad. The remaining subjects performed the reversed block order with a first dual-retrieval block, a second keypress single-retrieval block, and a third vocal single-retrieval block. Single-retrieval blocks and trials were identical to those of the single-retrieval criterion phase, with the exception that the experimenter additionally coded whether the voice key correctly registered the vocal response onset time for the vocal task. Dual-retrieval trials were identical to single-retrieval trials with the following exceptions. Subjects were instructed to speak both the digit and press the key as quickly as possible while being accurate. The cue remained on the screen until the subject executed both responses. Error feedback on dual-retrieval trials was presented for 3,500 ms. Half of the subjects were presented the seven cues at the even positions of the list presented in Table 1 while the remaining subjects were presented the seven cues at the odd positions of this list during the single-dual practice phase. Each of seven cues was presented once per block in randomized order.

Following the single-dual practice phase, subjects completed a *single-dual transfer phase* involving five triads (Transfer triads 1 – 5). Triads in this phase were identical to triads in the practice phase, the only exception being that the alternate set of seven cues in Table 1 was presented exclusively. We assume that the performance assessment in 5 triads is sufficient to test transfer effects (see below).

The experiment was conducted over three sessions. Session 1 involved the study and criterion phases, as well as the first 5 single-retrieval blocks of each task. Sessions 2 involved five additional single-retrieval blocks of each task and the first 10 triads of the single-dual

practice phase. Ten additional triads of the single-dual practice phase were performed to begin Session 3, followed by the five triads of the final single-dual transfer phase. Half of the subjects practiced the cue-response pairings of Condition 1 in Table 1, while the remaining subjects practiced cue-response pairings on Condition 2. These counter-balanced factors of task order and cue-response were crossed, with seven subjects in each of the four resulting cells.

Insert Table 2 here.

Results and Discussion

Raw data analyses

Accuracy and Voice-key results. In all analyses, we excluded either trials on which the voice key was tripped inappropriately based on the experimenter's judgment (typically, in these trials, the voice key was activated before a clear execution of a digit response or the key was not activated on a first digit response and required a second one) and/ or trials with RTs below 200 ms and above 5,000 ms (1.9 % of all single- and dual-retrieval trials). In the single-dual practice phase, error rates decreased from 3.6 % to 0.0 % for the keypress single-retrieval trials, and from 2.4 % to 1.2 % for the vocal single-retrieval trials. For the keypress task in the dual-retrieval blocks of this phase, error rates decreased from 1.9 % to 0.6 %, and for the vocal task from 6.0 % to 0.6 %. Dual-retrieval error rates during the single-dual practice phase decreased from 3.1 % to 0.6 % for the first completed response (keypress or vocal), and from 4.2 % to 0.6 % for the second response.

RT results. RTs averaged over all subjects for correctly performed single-retrieval (i.e., keypress task, vocal task) and for dual-retrieval trials (RT1, RT2) are shown in Figure 1**.** RTs decreased steadily over the course of single-dual practice (Practice triad 1 – 20), but increased markedly on the transfer test, particularly for dual-retrieval trials (Transfer triad 1).

11

Insert Figure 1 here.

Dual-retrieval results and model fits separated by response grouping mode

Identification of grouper and non-grouper subjects. Following Nino and Rickard (2003), we computed the mean IRI on dual-retrieval trials for each subject, averaging over all practice phase triads. The results are shown in Figure 2, individually ordered by IRI magnitude. Nino and Rickard observed a gap in their IRI plot at 300 ms and a similar but smaller gap is also present in the figure. Following their lead, we treat subjects with mean IRIs of less than 300 ms as response groupers and subjects with IRIs greater than 300 ms as non-groupers. Note that we are not necessarily suggesting that the 300 ms classification rule will perfectly capture the grouper vs. non-grouper distinction. Rather, that rule works well empirically as a basis for dividing subjects into sub-groups that have highly distinct performance characteristics and that facilitate theoretical interpretation.

On the basis of the 300 ms rule, 13 subjects were categorized as grouper subjects and 11 subjects were categorized as non-grouper subjects. The mean IRI and standard deviation among grouper subjects were 144 ms and 61 ms and among non-grouper subjects 639 ms and 211 ms, respectively.

During practice, these two groups of subjects did not differ with regard to single-vocal and single-keypress RTs and errors. In mixed measures ANOVAs including group (grouper subjects vs. non-grouper subjects) and triad (Triad 1 to Triad 20) on these data, this was evident from non-significant main effects of and interactions with group, Fs$(1, 22) < 1.421$, ps $> .25$, and, Fs$(19, 418) < 1.0$, respectively. Thus, there is no evidence that the categorization into groups based on IRIs is related to the performance levels on single-retrieval trials.

Insert Figure 2 here.

Practice phase RTs. Figure 3 shows the practice and transfer phase means for RT1 and RT2, along with the race and ES predictions, for non-grouper (Panel A) and grouper (Panel B) subjects in each triad. For non-grouper subjects, both RT1 and RT2 were clearly above the ES prediction on the first few triads, but tracked the ES predictions closely for the remaining triads, confirming the results of Nino and Rickard (2003). Dual-retrieval learning for non-grouper subjects thus appears to take the form of sequential retrievals carried out with increasing efficiency. That increased efficiency may involve a faster choice of which response to retrieve first, a shorter switch time between the two retrievals (Maquestiaux, Hartley, & Bertsch, 2004), or more efficient parallel processing of the motor stage for the first executed task with the retrieval stage for the second executed task (Sigman & Dehaene, 2006). There is no evidence in the mean RTs, however, for learned parallelism within the retrieval stage of processing for those subjects.

Given its success in accounting performance of non-grouper subjects, the ES model provides an empirically validated reference prediction for the evaluation of learned retrieval-stage parallelism among grouper subjects.[1] For those subjects, RT1 and RT2 were above or roughly equivalent to the ES prediction on the first two practice triads. With further practice, however, the mean RT2 for the grouper subjects fell well below the ES prediction, approaching the race RT2 prediction by the end of practice. A t-test comparing RT2 for grouper subjects to the ES prediction, in the last practice triad (Practice triad 20), was highly significant, $t(12) = 4.675$, $p < .001$, while there was no significant difference between RT2 and its race prediction, $t(12) = 1.229$, $p > .24$. These results clearly point to some form of learned retrieval parallelism for grouper subjects, again confirming Nino and Rickard (2003). RT1 for grouper subjects remained above the ES prediction throughout the single-dual

practice phase, as would be expected given our assumption that those subjects tended to delay execution of the first response until the second response had been retrieved.

To explore the results for the first practice triad further, we plotted cumulative distributions for RT2 during the first dual-retrieval practice block in Practice triad 1, alongside its respective race and ES predictions. The seven RTs (separately for RT2 and race/ ES predictions) for each subject were rank ordered from shortest to longest. Inclusion of all responses, regardless of accuracy, was necessary to maintain a valid RT ordering for each subject. Given the processing stage assumptions of the ES and race models, and the comparison based on two subsets of the data (single and dual), inclusion of error trials for both the observed dual-retrieval and the model predictions does not introduce bias. These RT distribution analyses extended the results for the means, showing that RT2 on the first dual-retrieval block was systematically longer than the ES prediction, even for grouper subjects (Figure 4). Thus, there is no evidence that the mode of grouped of response execution in itself leads to parallel retrieval. Rather, it is the combination of grouped response execution and dual retrieval practice that gives rise to learned parallelism.

At the end of practice (i.e., Practice triad 20), the cumulative distributions again confirmed the conclusions that RT2 for non-grouper subjects did not violate the ES prediction at any quantile in the distribution (Figure 4A). For grouper subjects (Figure 4B), however, RT2 compellingly violated the ES boundary prediction, confirming some form of learned retrieval parallelism throughout the distribution.

Insert Figure 3 & 4 here.

Transfer phase RTs. As evident in Panel A of Figure 3, RT data for non-grouper subjects were consistent with the ES prediction throughout the transfer phase, with the exception of slightly longer RTs on the first triad. Thus, the increased efficiency of sequential

retrieval that occurred during practice appears to have transferred, at least in part, to new cues, suggesting a task-level coordinative learning similar to that previously observed in other tasks (e.g., Liepelt et al., 2011, for the case of choice RT tasks). This conclusion is viable in light of the non-significant differences between RT1 and the corresponding ES prediction in the last practice triad (Practice triad 20) and the first transfer triad (Transfer triad 1), $F(1, 10) = 2.586$, $p > .13$, as well as similar differences between RT2 and the corresponding ES prediction in these triads, $F(1, 10) < 1.0$.

Of most interest in the transfer data is performance for grouper subjects on the first dual-retrieval transfer triad (Figure 3B). This first transfer triad provides optimal conditions to assess performance on new dual-retrieval cues because there were no prior dual-retrieval trials for those cues and hence no prior response chunking opportunity (see also Hazeltine, Aparicio, Weinstein & Ivry, 2007; Nino & Rickard, 2003). There was no evidence of learned parallelism on the first transfer triad, wherein RT2 was not significantly different from the ES prediction, $F(1, 12) < 1.0$. For grouper subjects we also observed a large increase in RT1 from the last practice triad to the first transfer triad. We interpret this effect in terms of continued use of the response grouping mode in the transfer phase, but in the context of a reversion back to sequential retrieval processing of two responses.

Cumulative distribution analyses for the first dual-retrieval block of the transfer phase (Transfer triad 1) are shown in Figure 4. For non-grouper and grouper subjects, RT2 is statistically equivalent to the ES prediction for all quantiles (non-groupers: ts$(10) < 2.079$, ps $> .07$; groupers, ts$(12) < 1$). Thus, the ES model is highly consistent with the RT2 results for both non-grouper and grouper subjects.

Finally, we explored, among the grouper subjects, whether the rate of dual-retrieval learning over the five transfer triads was different from that over the first five practice triads. The ANOVAs included factors of phase (practice vs. transfer) and triad (Triad 1 through Triad 5), performed on the RT2 data. We found no interaction between phase and triad, $F(4,$

48) < 1.0. Thus, dual-retrieval practice with one cue set (i.e., old cues) resulted in no observed improvement in the rate of dual-retrieval learning with an alternative cue set (i.e., new cues). Rather, the dual-retrieval learning rate appears to be entirely cue-specific.

Experiment 2

Results of grouper subjects in Experiment 1 are consistent with the cue-level chunking account of learned parallelism and they allow us to reject the purely task-level account as described in the Introduction. A variant of the task-level account, however, remains plausible: Grouper subjects may have made a shift to parallel retrieval during dual-retrieval practice but then shifted back to sequential retrieval on the transfer test in the context of new dual-retrieval cues and the absence of old dual-retrieval cues. That is, learned dual-retrieval parallelism may in principle be a task-level phenomenon, but whether or not the expression of that learning is observed may depend on its context. We will refer to this account as *context-dependent task-level account*.

In Experiment 2, we again investigated the cue-specificity of learned dual-retrieval parallelism and tested the cue-level chunking account versus the context-dependent task-level account. The primary change in design was that, on the transfer test, both old dual-retrieval cues and new dual-retrieval cues were randomly mixed. The cue-level account predicts that, on at least the first transfer triad, there will be a violation of the RT2's ES lower bound prediction among grouper subjects for old dual-retrieval cues (just as during the dual-retrieval practice phase) but no such violation for the new dual-retrieval cues. In contrast, the context-dependent task-level account assumes that sequential vs. parallel retrieval should apply to all cues of the current context, leading to one of two possible outcomes: 1) dual-retrieval for both old and new cues will be sequential and their RT2s will be consistent with the ES prediction, or 2) dual-retrieval for both old and new cues will be parallel and their RT2s violate the ES lower bound prediction.

Methods

Subjects

Twenty-four undergraduates at the University of California, San Diego, participated in the experiment for partial fulfillment of a course requirement.

Apparatus, Stimuli, Procedure, and Design were identical to Experiment 1 with the exception of presenting all 14 color word cues (practiced during the study, single-retrieval criterion, and single practice phase, see Table 1 & 2) in the vocal/ keypress single-retrieval blocks and the dual-retrieval blocks of the single-dual transfer phase.

Results and Discussion

Raw data analyses

Accuracy and Voice-key results. In all analyses, we excluded trials on which the voice key was tripped inappropriately based on the experimenter's judgment and/ or trials with RTs below 200 ms and above 5,000 ms (3.7 % of all single- and dual-retrieval trials). In the single-dual practice phase, error rates decreased from 1.8 % to 0.0 % for the keypress single-retrieval trials, and from 3.0 % to 1.2 % for the vocal single-retrieval trials. For the keypress task in the dual-retrieval blocks of this phase, error rates decreased from 3.9 % to 0.0 %, and for the vocal task from 8.3 % to 1.2 %. Dual-retrieval error rates during the single-dual practice phase, described with respect to which retrieval was completed first and which retrieval was completed second, decreased from 5.3 % to 0.0 % for the first response, and from 7.7 % to 0.6 % for the second response.

RT results. Single-dual practice and transfer phase RTs, averaged over all subjects for correctly performed single-retrieval (i.e., keypress task, vocal task) and dual-retrieval trials (RT1, RT2) are shown in Figure 5. Similarly to Experiment 1, RTs decreased steadily over the course of single-dual practice (Practice triad 1 – 20). On the transfer test, however, dual-retrieval RTs for new cues were markedly larger than those for old cues.

Insert Figure 5 here.

Dual-retrieval results and model fits separated by response grouping mode

Identification of grouper and non-grouper subjects. As in Experiment 1, we computed

the mean IRI on dual-retrieval trials for each subject, averaging over all practice phase triads.

The results are shown in Figure 6, ordered by IRI magnitude. As observed in the previous

studies, there is a gap in the distribution of IRIs around 300 ms that suggests qualitative

strategy differences, i.e. grouping vs. non-grouping as described earlier. On this basis, 17

subjects were categorized as grouper subjects and seven subjects were categorized as non-

grouper subjects. The mean IRI and standard deviation among grouper subjects were 117 ms

and 47 ms and among non-grouper subjects 599 ms and 164 ms, respectively.

During practice, these two groups of subjects did not differ with regard to single-vocal

and single-keypress RTs and errors. In mixed measures ANOVAs including group (grouper

subjects vs. non-grouper subjects) and triad (Triad 1 to Triad 20) on these data, this was

evident from non-significant main effects of and interactions with group, $Fs(19, 418) < 1.0$,

and, $Fs(19, 418) < 1.0$, respectively. Thus, similar to Experiment 1, there is no evidence that

categorization into groups based on IRIs is related to the performance levels on single-

retrieval trials.

Insert Figure 6 here.

Practice phase RTs. Figure 7 shows the practice phase means for RT1 and RT2, along

with the race and ES predictions, for non-grouper (Panel A) and grouper (Panel B) subjects in

each triad. As in Experiment 1, for non-grouper subjects, both RT1 and RT2 were above the

ES prediction on the first few triads, but tracked the ES predictions closely for the remaining

triads. Dual-retrieval learning for non-grouper subjects thus appears to take the form of sequential retrievals carried out with increasing efficiency, involving faster choice of which response to retrieve first, a shorter switch time between the two retrievals (Maquestiaux et al., 2004), or more efficient parallel processing of the motor stage for the first executed task with the retrieval stage for the second executed task (Sigman & Dehaene, 2006). There is no evidence, however, for learned parallelism within the retrieval stage of processing for those subjects.

For grouper subjects, both RT1 and RT2 were above or roughly equivalent to the ES prediction on the first triad. With further practice, however, the mean RT2 for grouper subjects fell well below the ES prediction, approaching the race RT2 prediction by the end of practice. A t-test comparing RT2 for grouper subjects to the ES prediction, in the last practice triad (Practice triad 20), was highly significant, $t(16) = 5.216$, $p < .01$, whereas there was no significant difference between RTs and its race prediction, $t(16) = 1.125$, $p > .28$. These results, like those of Experiment 1, point to some form of learned retrieval parallelism for grouper subjects. As in Experiment 1, RT1 for grouper subjects remained above the ES prediction throughout the single-dual practice phase, as would be expected given our assumption that those subjects tended to delay execution of the first response until the second response had been retrieved.

Practice phase RT2 distribution results, derived identically to those of Experiment 1, are shown in Figure 8. For non-groupers, RT2 did not fall below the ES prediction for any quantiles on either the first or the 20th practice triad. For grouper subjects, RT2 was equivalent to the ES prediction on the first triad but fell systematically below it on the 20th triad. Thus, as for Experiment 1, there is (1) no evidence that the mode of grouping of motor responses potentially results from a violation of the ES at some point of the first triad's distribution and (2) a confirmation of some form of retrieval parallelism throughout the distribution.

Insert Figure 7 & 8 here.

Transfer phase RTs. As evident in Panel A of Figure 7 and similar to Experiment 1, non-grouper subjects' data were consistent with the ES prediction throughout the transfer phase, with the exception of slightly longer RT1 on the first transfer triad. As evidenced by similar differences between RT1 and the relative ES prediction in the last practice triad (Practice triad 20) and the first transfer triad (Transfer triad 1) as well as similar differences between RT2 and the relative ES prediction in these triads, $Fs(1, 6) < 1.0$, the present data provide no evidence for effects between these triads that are specific for dual-retrieval processing. This is consistent with the assumption of increased efficiency of sequential retrieval that occurred during practice and appears to have transferred to new cues, suggesting a task-level learning effect.

Of most interest in the transfer data is performance for grouper subjects on the first dual-retrieval transfer triad (Figure 7B). The interaction of dataset (RT2 vs. ES prediction) and cue type (old cues vs. new cues) was significant, $F(1, 16) = 4.772$, $p < .05$. Whereas RT2 for old cues violated the ES lower bound prediction on the first transfer triad, $t(16) = 2.185$, $p < .05$, that violation was not observed for new cues, $t(16) < 1$. These findings are consistent with the assumptions of learned retrieval parallelism at the cue level rather than at the context level. Also for grouper subjects, RT1 for new cues on the first transfer block was far above the ES prediction and, indeed, approached the RT2 values. As for Experiment 1, we interpret this effect as reflecting continued use of the response grouping mode in the transfer phase, but in the context of a reversion back to sequential retrieval stage processing, as in Experiment 1.

Cumulative distribution analyses for the first dual-retrieval block of the transfer phase are shown in Figure 9. For both non-grouper (Panel A) and grouper subjects (Panel B), RT2 for new cues is statistically equivalent to the ES prediction for all quantiles (non-groupers:

ts(6) < 1.530, ps > .18; groupers: ts(16) < 1.908, ps > .08). In contrast, the lower RT2 quantiles for grouper subjects violate the ES prediction for old cues. In detail, the interaction of quantile (Quantile 1 – 7) and data set (observed vs. ES prediction) on RT2 data was significant, F(6, 78) = 3.739, p < .01, with observed RT2s being significantly below the ES predictions in each of the lower five quantiles, ts(16) > 3.47, ps < .01. In summary, the Experiment 2 results are consistent with the assumption of cue-specific learned parallelism in the case of two retrievals from a single cue.

Insert Figure 9 here.

General Discussion

We examined three hypothesis about the specificity of learned retrieval parallelism that is observed for subjects who execute their responses in close temporal proximity (i.e. grouper subjects): 1) learned parallelism is a task-level (i.e., task-general) phenomenon, 2) learned parallelism can generalize to new cues, but only in the context of old cues (the context-dependent task-level account), and 3) learned parallelism is specific (i.e., limited) to each practiced cue. In sum, our data provide evidence against both the task-level and context-dependent accounts in favor of the cue-specific account. In Experiment 1, learned parallelism for grouper subjects did not transfer to a set of new dual-retrieval cues that had previously been trained only on single retrieval trails, eliminating the purely task-level account. In Experiment 2, learned parallelism did not transfer to new cues when mixed with the previously practiced old cues for which some form of learned parallelism already had been achieved. These findings rule out the possibility that parallel retrieval generalizes to new cues in a context in which parallel retrieval is occurring for old cues. Instead, grouper subjects in Experiment 2 continued to retrieve responses in parallel for old cues (i.e., dual-retrieval RT2 violated the ES prediction) but appear to have shifted back to sequential retrieval for new cues

(i.e., dual-retrieval RT2 did not violate its ES prediction), results that are fully consistent with cue-specific learned parallelism. These results held for both mean RTs and the RT distribution. The observed cue specificity of learned parallelism for grouper subjects is consistent with the retrieval model suggested by Nino and Rickard (2003). In their model, there is (1) a structural memory retrieval bottleneck and (2) a cue-specific response chunking mechanism that allows both retrievals to be performed with one pass through the bottleneck.

One alternative to Nino and Rickard's (2003) account that has not yet been considered is that subjects make a parallel vs. sequential strategic choice on each dual-retrieval trial of the transfer test. That account cannot strictly be eliminated, but it encounters several unique problems that render it unlikely. First, that account would require that, when presented with a cue for dual retrieval, subjects in Experiment 2 first evaluate whether the cue is old or new and then make a choice to proceed with either parallel or sequential retrieval. That decision would have to be made independently on every dual-retrieval trial, adding substantially to the cognitive effort required to perform dual-retrieval tasks. Such inefficient sequential retrieval would be expected to yield RT2s for new cues that are clearly above the ES model prediction. However, both the grouper and non-grouper subjects' data are inconsistent with this assumption and statistically show similar RT2 and ES model predictions (Experiment 1: grouper subjects, $t(12) < 1$/ non-grouper subjects, $t(10) < 1$; Experiment 2: grouper subjects, $t(16) < 1$/ non-grouper subjects, $t(6) < 1$), demonstrating the efficiency of sequential retrieval with no indicator of effortful trial-by-trial decision on serial vs. parallel strategy decision.

Second, it is not clear what information subjects would use to make that trial-level strategic decision. They might rely on cue familiarity (Farley & Keating, 2009) or they might use information gained from an early read of retrieval fluency (e.g., Kim, Park, & Wagner, in press) to discriminate old from new cues. However, given the substantial overlap in the single-retrieval RTs for old and new cues in Experiment 2's transfer phase, as illustrated in Appendix B, it is unlikely that judgements fluency, retrieval difficulty or prior response time

would allow subjects to reliably and efficiently separate all old cues from new cues. As such, under the strategic model, subjects would end up performing parallel retrieval for a subset of both old and new cues. The expected result would be a crossover effect between RT2 and the ES prediction in the cumulative distribution plots for both old and new cues. That effect, however, was not observed. Third, data from other studies demonstrates that subjects, at the end of practice, consistently chose one type of processing strategy when strategy reports are collected (Rickard, 1997, 2004; Touron & Hertzog, 2004; Touron, Hoyer, & Cerella, 2004). For instance, there is evidence for the report of memory-retrieval processing of arithmetic problems on nearly all trials with sufficient practice (Rickard, 1997). Similar findings of a practice-dependent uniform strategy use are reported from the context of the noun pair lock-up task (Touron & Hertzog, 2004). Finally, if a strategic bottleneck is defined such that it can allow for either parallel or sequential retrieval at any and all levels of specificity, with no predictive constraints on that flexibility, then that model may be difficult if not impossible to falsify. In contrast, the structural bottleneck plus response chunking model is strongly constrained in important respects and clearly excludes outcomes that were plausible a priori.

Nino and Rickard (2003) offer two specific accounts of cue-level response chunking in the context of a structural bottleneck model. According to the first account, which we advanced in the introduction, response chunking occurs due to the creation and use of a compound representation of the cue and task sets (i.e., retrieve digit and retrieve keypress) from which associations to both responses are formed after sufficient practice. The structural bottleneck is preserved in this account because only one compound representation is needed to mediate both responses on any given dual-retrieval trial. Alternatively, provided that participants group their responses, dual-retrieval practice may form a direct, chained association between two responses, strengthened to the point at which retrieval of the first response may serve as a cue that leads directly to priming and retrieval of the second response. Because this chained association account does not require restarting the second

retrieval from the originally presented cue and instead relies on an association between two response systems, it is possible that performance via this mechanism may become faster than predicted by the ES model following practice, as was observed for grouper subjects. There are two arguments, however, that identify a compound retrieval account as more plausible then the chained association account. First, it does not seem plausible that subjects could execute associative retrievals as fast as demonstrated by some grouper subjects, for whom the RT2 distribution approached the race model prediction and IRIs were below 100 ms. Second, the chained association account could predict that grouper subjects are highly consistent in their response order for a given cue, given that the association would presumably operate in only one direction (keypress to vocal digit or the reverse) for efficient parallel retrieval. To compute this *response-order consistency*, we focused on dual-retrieval blocks of the last practice phase (i.e., Triad 16 - 20) in grouper and non-grouper subjects in Experiment 1 and 2. For each individual cue (i.e., color word), we computed the variability in response order during these last practice triads. This computation results in a standard deviation value for each cue and each participant. The comparison of these aggregated values showed a lower consistency (higher variability) for grouper than non-grouper subjects in Experiment 1, $t(22) = 4.050$, $p < .001$, and Experiment 2, $t(22) = 2.999$, $p < .01$ .Overall, the data are more consistent with the compound representation account than the chained association account.

From a broader perspective, the cue-specific chunked responses account might explain effects of configural motor response learning (Hazeltine et al., 2007). In that study, subjects performed configurations of three simultaneous, piano-chord like keypresses. In a final transfer phase, executions of these practiced (i.e., old) configurations were faster than new, unpracticed configurations that included old, practiced keypresses. That study thus suggests the retrieval of one practiced configuration (i.e., chunked retrieval) can occur simultaneously for all keypresses, whereas for unpracticed configurations there is no such simultaneous (i.e., chunked) retrieval of keypress information. Those findings mirror the present findings for

grouper subjects: There is parallel retrieval on old, practiced cues but no parallel retrieval on new, unpracticed cues. In contrast, there exist findings from dual-task practice situations that might not be consistent with chunking in the context of Nino and Rickard's (2003) bottleneck model. In situations involving two-choice RT tasks (Schumacher et al., 2001, Strobach, Frensch, Müller, & Schubert, 2012c, 2012d), for example, exclusive single-task practice on some cues results in final dual-task performance that is as good as that achieved through mixed dual- and single-task practice on other cues (Hazeltine et al., 2002). The authors interpreted their results to reflect parallel and independent central response-selection stages for the two tasks. However, those findings can also be explained with latent (structural) bottleneck processing: Response selection stages (i.e., the presumed bottleneck stages in choice RT tasks) are extremely shortened and scheduled such that there is no temporal overlap, and thus no interference, between these stages (Anderson et al., 2005; Ruthruff et al., 2003; Schubert, 2008; however see Ruthruff et al., 2006). In that account, the two tasks are processed independently and unchunked, but the structural bottleneck characteristic remains (cf. Nino & Rickard, 2003).

One important difference between cases of dual-choice RT tasks and the present dual-retrieval tasks may explain why our findings supported chunked processing while Hazeltine et al.'s (2002) do not. In our task, there was only one cue with the responses, whereas in their task there were two cues and one response for each cue. Presentation of separate cues may promote separate processing of two tasks, whereas a single cue may promote response chunking. Further studies are needed to test this possibility systematically, using the cases of dual-retrieval and dual-choice RT tasks that involve cases of one cue vs. two cues (see also Fagot & Pashler, 1992).

Finally, the data for non-grouper subjects indicates that coordination of sequential retrieval becomes more efficient over the course of practice, allowing RTs to converge toward, but not fall below, a lower bound sequential retrieval prediction embodied in the ES.

This result extends prior work showing that dual-task choice RT practice can improve coordinative aspects of performance (Kamienkowski, Pashler, Sigman, & Dehaene, 2011; Kramer et al., 1995; Strobach et al., 2012b). For non-grouper subjects in the current experiments, both RT1 and RT2 tended to converge on the ES prediction following dual-retrieval practice. That finding suggests that dual-retrieval practice can result in nearly optimal task scheduling and coordination of perceptual, central (retrieval), and motor processing stages. As observed here and by Liepelt et al. (2011) for dual-choice RT tasks, that coordinative learning may transfer to new cues and tasks, respectively, potentially representing general-purpose control skills. In line with such skills, the response execution mode (grouping vs. non-grouping) also transferred to new cues on the transfer test in the present experiments, as indicated by the similar performance patterns at the beginning of the single-dual practice phase and the transfer phase, within both sets of grouper and non-grouper subjects.

Promising future experiments might focus on whether there are any conditions at all under which parallel retrieval is possible prior to dual-retrieval practice. One possibility might involve extensive single-retrieval practice in the context of a transfer test (such as that of Experiment 2) to both old and new dual-retrieval cues. In such a design, the total amount of practice might also be controlled such that both old and new cues are countered with equal frequency during practice. We are skeptical, however, that those experimental conditions would qualitatively change the outcome for new dual-retrieval cues on the transfer test. Nino and Rickard (2003, Experiment 1) have already demonstrated that extensive single-cue practice is not sufficient to yield parallel dual retrieval. It seems unlikely that, in light of the current results, adding both old and new cues to the transfer test would change that outcome.

Further empirical work might focus on the underlying basis for the individual differences in response grouping. This basis could be either strategic and sensitive to instructional manipulation or a more fundamental and non-malleable individual difference.

26

Other potential avenues could include exploration of dual-retrieval effects across life-time developmental and across different memory systems and modes of retrieval.

Conclusions

The empirical results indicate that (1) subjects who do not group responses on dual-retrieval trials do not achieve learned retrieval parallelism, but they can achieve nearly optimal sequential retrieval performance via dual-retrieval practice, (2) this improved efficiency for non-grouper subjects transfers at least partially to new dual-retrieval cues, (3) for subjects who do group responses, learned retrieval parallelism does occur, 4) the mechanism of learned parallelism for grouper subjects is cue-specific response chunking, which may be facilitated by the concurrent response activation that is a consequence of the grouped response strategy. Result 4 is consistent with models that assume a structural retrieval bottleneck along with a cue-specific response chunking mechanism that allows both responses to be retrieved in one pass through that bottleneck.

References

Anderson, J. R., Taatgen, N. A., & Byrne, M. D. (2005). Learning to achieve perfect time sharing: Architectural implications of Hazeltine, Teague, & Ivry (2002). *Journal of Experimental Psychology: Human Perception and Performance, 31,* 749–761.

Fagot, C., & Pashler, H. (1992). Making two responses to a single object: Implications for the central attentional bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1058-1079.

Farley, A. P., & Keating, G. D. (2009). Familiarity effects on lexical access during L2 word reading. In C. Brantmeier (Ed.) , *Crossing languages and research methods: Analyses of adult foreign language reading* (pp. 31-46). Charlotte, NC US: IAP Information Age Publishing.

Hazeltine, E., Aparicio, P., Weinstein, A., & Ivry, R. B. (2007). Configural response learning: The acquisition of a nonpredictive motor skill. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1451-1467.

Hazeltine, E., Teague, D., & Ivry, R. B. (2002). Simultaneous dual-task performance reveals parallel response selection after practice. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 527-545.

Hebb, D. O. (1949). *The Organization of Behavior: A neuropsychological theory*. New York: Wiley.

Hirst, W., Spelke, E. S., Reaves, C. C., Caharack, G., & Neisser, U. (1980). Dividing attention without alteration or automaticity. *Journal of Experimental Psychology: General, 109*, 98–117.

Jaeggi, S.M., Buschkuehl, M., Jonides, J., and Perrig, W.J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences* 105**,** 6829-6833.

Kamienkowski, J. E., Pashler, H., Sigman, M., & Dehaene, S. (2011). Effects of practice on task architecture: Combined evidence from interference experiments and random-walk models of decision making. *Cognition*, 119, 81-95.

Kim, Y., Park, C., & Wagner, R. K. (in press). Is oral/text reading fluency a "bridge" to reading comprehension? *Reading And Writing*,

Kramer, A. F., Larish, J. F., & Strayer, D. L. (1995). Training for Attentional Control in Dual Task Settings: A Comparison of Young and Old Adults. *Journal of Experimental Psychology: Applied, 1*(10), 50-76.

Li, Y., & Wright, D. L. (2000). An assessment of the attention demands during random and blocked practice schedules. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53*(A), 591–606.

Liepelt, R., Strobach, T., Frensch, P. A., & Schubert, T. (2011). Improved inter-task coordination after extensive dual-task practice. *Quarterly Journal of Experimental Psychology. 64*(7), 1251–1272.

Logan, G. D., & Gordon, R. (2001). Executive control of visual attention in dual-task situations. *Psychological Review*, *108*(2), 393-434.

Maquestiaux, F., Hartley, A. A., & Bertsch, J. (2004). Can practice overcome age-related differences in the psychological refractory period effect? *Psychology and Aging, 19*, 649-667.

Meyer, D. E., & Kieras, D. E. (1997a). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104*, 3-65.

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

Nino, R. S., & Rickard, T. C. (2003). Practice effects on two memory retrievals from a single cue. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(3), 373-388.

Oberauer, K., & Bialkova, S. (2011). Serial and parallel processes in working memory after practice. *Journal Of Experimental Psychology: Human Perception And Performance*, *37*(2), 606-614.

Oberauer, K., & Kliegl, R. (2004). Simultaneous cognitive operations in working memory after dual-task practice. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 689-707.

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

Pashler, H. (1998). *The psychology of attention.* Cambridge, MA: MIT Press.

Rickard, T. C. (2004). Strategy execution in cognitive skill learning: an item-level test of candidate models. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*(1), 65-82.

Rickard, T. C., & Pashler, H. (2005). A Bottleneck in Memory Retrieval from a Single Cue. *Center for Human Information Processing Technical Report 2005-01.*

Ruthruff, E., Johnston, J. C., Van Selst, M. V., Whitsell, S., & Remington, R. (2003). Vanishing dual-task interference after practice: Has the bottleneck been eliminated or is it merely latent? *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 280-289.

Ruthruff, E., Pashler, H. E., & Klaassen, A. (2001). Processing bottlenecks in dual-task performance: Structural limitation or strategic postponement? *Psychonomic Bulletin & Review*, *8*(1), 73-80.

Ruthruff, E., Van Selst, M., Johnston, J. C., & Remington, R. W. (2006). How does practice reduce dual-task interference: Integration, automatization, or simply stage-shortening? *Psychological Research, 70*, 125-142.

Schubert, T. (1999). Processing differences between simple and choice reaction affect bottleneck localization in overlapping tasks. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 408–425.

Schubert, T. (2008). The central attentional limitation and executive control. *Frontiers of Bisocience*, *13*, 3569-3580.

Schumacher, E. H., Seymour, T. L., Glass, J. M., Fencsik, D. E., Lauber, E. J., Kieras, D. E., & Meyer, D. E. (2001). Virtually perfect time sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological Science, 12*(2), 101-108.

Sigman, M., & Dehaene, S. (2006). Dynamics of the central bottleneck: Dual-task and task uncertainty. *PLOS: Biology*, *4*, e220.

Smith, M.A., Roediger, H.L., & Karpicke, J.D. (in press). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

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

Strobach, T., Frensch, P. A., Müller, H. & Schubert, T. (2012c). Age- and practice-related influences on dual-task costs and compensation mechanisms under optimal conditions for dual-task performance. *Aging, Neuropsychology, and Cognition, 19*(1-2), 222-247.

Strobach, T., Frensch, P. A., Müller, H. & Schubert, T. (2012d). Testing the limits of optimizing dual-task performance in younger and older adults. *Frontiers in Human Neuroscience*, *6*:39.

Strobach, T., Frensch, P. A., Soutschek, A., & Schubert, T. (2012b). Investigation on the improvement and transfer of dual-task coordination skills. *Psychological Research*, *76*(6), 794-811.

Strobach, T., Liepelt, R., Schubert, T. & Kiesel, A. (2012a). Task switching: effects of practice on switch and mixing costs. *Psychological Research, 76*(1), 74-83.

Tombu, M., & Jolicoeur, P. (2004). Virtually no evidence for virtually perfect time-sharing. *Journal of Experimental Psychology: Human Perception & Performance, 30*, 795-810.

Touron, D. R., & Hertzog, C. (2004). Strategy shift affordance and strategy choice in young and older adults. *Memory & Cognition*, *32*(2), 298-310.

Touron, D. R., Hoyer, W. J., & Cerella, J. (2004). Cognitive Skill Learning: Age-Related Differences in Strategy Shifts and Speed of Component Operations. *Psychology And Aging*, *19*(4), 565-580.

Townsend, J. T., & Wenger, M. J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review*, *11*(3), 391-418.

Van Selst, M., Ruthruff, E., & Johnston, J. C. (1999). Can practice eliminate the Psychological Refractory Period Effect? *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1268-1283.

Watson, J. M., & Strayer, D. L. (2010). Supertaskers: Profiles in extraordinary multitasking ability. *Psychonomic Bulletin & Review*, *17*(4), 479-485.

Appendix

Appendix A

Quantitative derivation of the efficient sequential (ES) and the race model predictions. The quantitative ES and race models described here are closely related to those described by Nino and Rickard (2003). Both models assume sequential and stochastically independent (e.g., Sternberg, 1969) perceptual, memory retrieval, and motor stages of processing (*p, r,* and *m,* respectively). On a given single-retrieval trial, both models assume that,

$$RTk = Lp + Lrk + Lmk \quad \text{(for the keypress task)}, \quad (1a)$$

$$RTv = Lp + Lrv + Lmv \quad \text{(for the vocal task)}, \quad (1b)$$

where Lp is random variate from the latency distribution of the perceptual stage for a given cue (and subject), Lrk (or Lrv) is a random variate from the latency distribution of the keypress (or vocal-digit) retrieval stage for a given cue, and Lmk (or Lmv) is a random variate from the latency distribution for the motor stage for a given response.

The ES model of dual-retrieval. The ES model assumes a bottleneck at the retrieval stage that allows for processing of only one retrieval at-a-time. However, it assumes that there is no bottleneck or capacity constraint on parallel execution of a retrieval stage and a motor stage, or of a keypress motor stage and a vocal motor stage. Assuming for now that the motor stage is always executed as soon as the retrieval stage is complete (i.e., the presumed case for subjects identified as non-grouper subjects), the ES model predicts that RT1 on a given dual-retrieval trial will involve that same stage sequence as does a single-retrieval trial:

$$RT1 = Lp + Lrk + Lmk \quad \text{(if the keypress response is executed first)},$$

$$\text{or,} \quad (2a)$$

$$RT1 = Lp + Lrv + Lmv \quad \text{(if the vocal response is executed first)}.$$

The ES additive stage prediction for RT2 is either:

$$RT2 = Lp + Lrk + Lrv + Lmv \quad \text{(if the keypress response is executed first)},$$

$$\text{or,} \quad (2b)$$

$$RT2 = Lp + Lrv + Lrk + Lmk \quad \text{(if the vocal response is executed first)},$$

where RT2 includes the latency components of the perceptual stage, the retrieval stages of both tasks , and the motor stage of the second executed task (i.e., keypress or vocal task).

Deriving the ES dual-retrieval predictions from the single-retrieval data. The ES model predicts that a random variate from the RT1 distribution on each dual-retrieval trial is estimated simply by the RT for the corresponding single-retrieval trial on the adjacent block (or more precisely, from blocks within the same triad as the target dual-retrieval trial; see pages 14 and 24 of the text). Thus, for each dual-retrieval trial for each subject, a matched ES random variate, either RTk or RTv, can be selected.

Similarly, for each RT2 value, a matched random variate can be computed from single-retrieval trials as:

$$RT2 = RTk + RTv - 200 \quad \text{(if the keypress response is completed first)}$$

or, \hfill (3b)

$$RT2 = RTk + RTv - 250 \quad \text{(if the vocal response is completed first)}.$$

To understand the rationale behind Equation 3b, compare the ES additive stage equation for RT2 for the example case in which the keypress retrieval stage is executed first,

$$RT2 = Lp + Lrk + Lrv + Lmv, \hfill (4)$$

to the additive stage equation if RT2 were predicted by the simple sum of the two single-retrieval latencies:

$$RT2 = Lp + Lrk + Lmk + Lp + Lrv + Lmv. \hfill (5)$$

Relative to the true ES prediction for RT2 (Equation 4), Equation 5 has an extra Lp term and an extra Lmk term. To jointly compensate for these extra perceptual and motor terms, Nino and Rickard (2003; see also Rickard & Pashler, 2005) used empirically motivated correction factors of -200 ms when vocal response is executed first and of -250 ms when the keypress response is executed first (see Nino & Rickard for further discussion of these correction factors), yielding Equation 3b for RT2.

The race model of dual-retrieval. Based on prior results, the race model is not expected to provide a satisfactory overall account of the data. It is included, however, as a reference prediction for highly efficient parallel retrieval. In deriving this model's predictions, we assume that a single perceptual event, Lp, runs to completion first, followed by a race between the combined retrieval and motor stages for the vocal retrieval (Lrv + Lmv) and the keypress retrieval (Lrk + Lmk). On a given dual-retrieval trials, then,

$$RT1 = Lp + \min [(Lrv + Lmv), (Lrk + Lmk)], \qquad (6a)$$

where min [(Lrv + Lmv), (Lrk + Lmk)] denotes the shorter of the two retrieval plus motor stage latencies on that trial. Analogously,

$$RT2 = Lp + \max [(Lrv + Lmv), (Lrk + Lmk)]. \qquad (6b)$$

Deriving the race dual-retrieval predictions from the single-retrieval data. Nino and Rickard (2003) showed that the "dangling" Lp term in Equations 6 a and b has negligible influence on the statistical race dynamics (because variance in the Lp stage can be assumed to be a small fraction of the overall variance), allowing the matched race random deviates, based on the single-retrieval trials, to be simplified to:

$$RT1 = \min [RTk, RTv] \qquad (7a)$$

$$RT2 = \max [RTk, RTv]. \qquad (7b)$$

General notes and assumptions. Using the procedures described above, each dual-retrieval trial for each subject has matched RT1 and RT2 predictions (i.e., theoretically matched random variates) corresponding to each model. Three data sets can thus be defined: 1) the observed dual-retrieval data set, 2) the trial-matched ES prediction data set based on single-retrieval trials, and 3) the trial-matched race prediction data set based on single-retrieval trials. In analyses of the mean RTs and the RT distribution quantiles, the observed dual-retrieval data sets were compared to these theoretical ES and race data sets.

The predictions for both the ES and race models should be understood as idealized predictions that correspond to best case performance, with no inefficiences in task execution

(see the main text for discussion of possible performance inefficiencies). Thus, the model predictions for mean RTs and for distribution quantiles should be understood as lower bound predictions. If either model is fundamentally on the right track in its core theoretical claims, and if dual-retrieval practice has the effect of reducing the latency impact of any initial performance inefficiencies, then dual-retrieval RTs would be expected to converge asymptotically on that model's predictions following dual-retrieval practice. On the other hand, if in any case the observed dual-retrieval RTs fall substantially below a model's predictions, then that model can be confidently rejected for that case in its current form.

Note that, in a departure from Nino and Rickard (2003), we excluded the 100 ms dual-retrieval trial preparation cost (Li & Wright, 2000; Pashler, 1998; Sigman & Dehaene, 2006) from the current model development. We excluded this cost because we aimed to provide idealized, lower bound RT predictions for both models. Furthermore, the inclusion of preparation costs in the Nino and Rickard derivation seem rather arbitrary because (1) there are no previous studies providing evidence for a preparation cost in the case of two retrievals from a single cue and (2) any preparation cost that may be associated with dual-retrieval in the current experiments may diminish substantially as a result of practice (e.g., Liepelt et al., 2011, for the case of choice RT dual-tasks), and it is performance after dual-retrieval practice that is of primary interest here.

Finally, two implicit assumptions in the derivation of dual-retrieval predictions based on the single-retrieval data should be noted. First, it is assumed that each subject's mental state, with respect to factors such as motivation, speed-accuracy bias, and readiness, was the same on average for trials in adjacent single- and dual-retrieval blocks. The data appear to be consistent with this assumption. There is no evidence, for example, of speed-accuracy trade-offs between adjacent single- and dual-retrieval blocks at the end of the single-dual practice phases or in the transfer phases. A second and related assumption is that any retrieval practice effects on RTs across adjacent blocks have only a negligible biasing effect on the model

predictions. Given the consistent graphical tracking that was observed over multiple blocks for the observed vs. ES prediction for dual-retrieval performance, and the counter-balancing over subjects of the order of single- vs. dual-retrieval blocks in Experiment 2, this assumption appears to be well justified.

Appendix B

Insert Figure 10 here.

Appendix B

Acknowledgments

Footnotes

Footnote 1. The success of the ES model in accounting performance of non-grouper subjects supports its use as a reference prediction for the evaluation of learned retrieval-stage parallelism among grouper subjects. Further, the success of that model renders alternative theoretical models, and in particular capacity sharing models, less compelling as candidate accounts. Whereas the ES model makes precise and testable claims about the lower bound for RT2 (prior to dual-retrieval practice) even in the case of response grouping, capacity based models do not, unless the free parameters of a capacity-based system are independently constrained.

Tables

Table 1. Cue-response mapping pairs for all cues.

	Condition 1			Condition 2	
Color words	vocal-digit response	Keypress response	Color words	vocal-digit response	Keypress response
Red	5	←	Red	8	←
Green	4	←	Green	5	→
Blue	1	→	Blue	6	→
Yellow	3	→	Yellow	4	→
Purple	2	←	Purple	0	→
Brown	6	→	Brown	9	←
Black	7	←	Black	2	←
Orange	8	←	Orange	1	←
White	9	→	White	3	←
Pink	0	→	Pink	7	→
Gold	12	←	Gold	11	→
Silver	11	←	Silver	14	←
Grey	14	→	Grey	13	→
Olive	13	←	Olive	12	←

Table 2. Overview of the general procedure in Experiment 1 and 2.

Session	Experimental phases	Number of blocks/ triads
1	Study phase	
	Single-retrieval criterion phase	
	Single-retrieval practice phase	10 blocks each task
2		
	Single-dual practice phase	20 triads
3		
	Single-dual transfer phase	5 triads

Figures captions

Figure 1. Observed reaction times (RTs) in single-retrieval blocks of the keypress task and the vocal task of Experiment 1; observed RTs in dual-retrieval blocks (i.e., RT1 and RT2) in the overall dataset during the 20 practice triads and 5 transfer triads.

Figure 2. Inter-response interval (IRI) of individual subjects of Experiment 1.

Figure 3. Observed reaction times (i.e., RT1 and RT2) as well the predictions of the race and the efficient sequential (ES) models in the non-grouper subjects (Panel A) and grouper subjects (Panel B) during the single-dual practice and transfer phases of Experiment 1.

Figure 4. Distribution of observed reaction times for slower responses (i.e., RT2) as well as their race and efficient sequential (ES) predictions in non-grouper subjects (A) and grouper subjects (B) divided into 7 quantiles in Practice triad 1 (start of practice), Practice triad 20 (end of practice), and Transfer triad 1 (start of transfer) of Experiment 1.

Figure 5. Observed reaction times (RTs) in single-retrieval blocks of the keypress task and the vocal task of Experiment 2; observed RTs in dual-retrieval blocks (i.e., RT1 and RT2) in the overall dataset during the 20 practice triads and 5 transfer triads. Increased RTs in Triad 11 reflect the start of Session 3 after the end of Session 2 (Triad 10).

Figure 6. Inter-response interval (IRI) of individual subjects of Experiment 2.

Figure 7. Observed reaction times (i.e., RT1 and RT2) as well the predictions of the race and the efficient sequential (ES) models in the non-grouper subjects (Panel A) and grouper subjects (Panel

B) during the single-dual practice and transfer phases of Experiment 2. Increased RTs in Triad 11 reflect the start of Session 3 after the end of Session 2 (Triad 10).

Figure 8. Distribution of observed reaction times for slower responses (i.e., RT2) as well as their race and efficient sequential (ES) predictions in non-grouper subjects (A) and grouper subjects (B) divided into 7 quantiles in Practice triad 1 (start of practice), Practice triad 20 (end of practice), and Transfer triad 1 (start of transfer) of Experiment 2.

Figure 9. Distribution of observed reaction times of second responses (i.e., RT2) as well as their race and efficient sequential (ES) predictions for old cues and new cues in (A) non-grouper subjects and (B) grouper subjects divided into 7 quantiles in Transfer triad 1 of Experiment 2. Asterisks (*) indicate significant differences between observed RT2 and its ES prediction at particular quantile levels.

Figure 10. Cumulative trial proportion on single-retrieval RTs in the keypress task and the vocal-digit task during Experiment 2' transfer phase. The graphs represent RTs on old and new cues in bins of 250 ms (i.e., Bin 1: RTs 0 – 250 ms, Bin 2: RTs 250 – 500 ms, …). Inclusion of all responses, regardless of accuracy, was necessary to maintain a valid RT ordering for each subject in these distribution analyses.
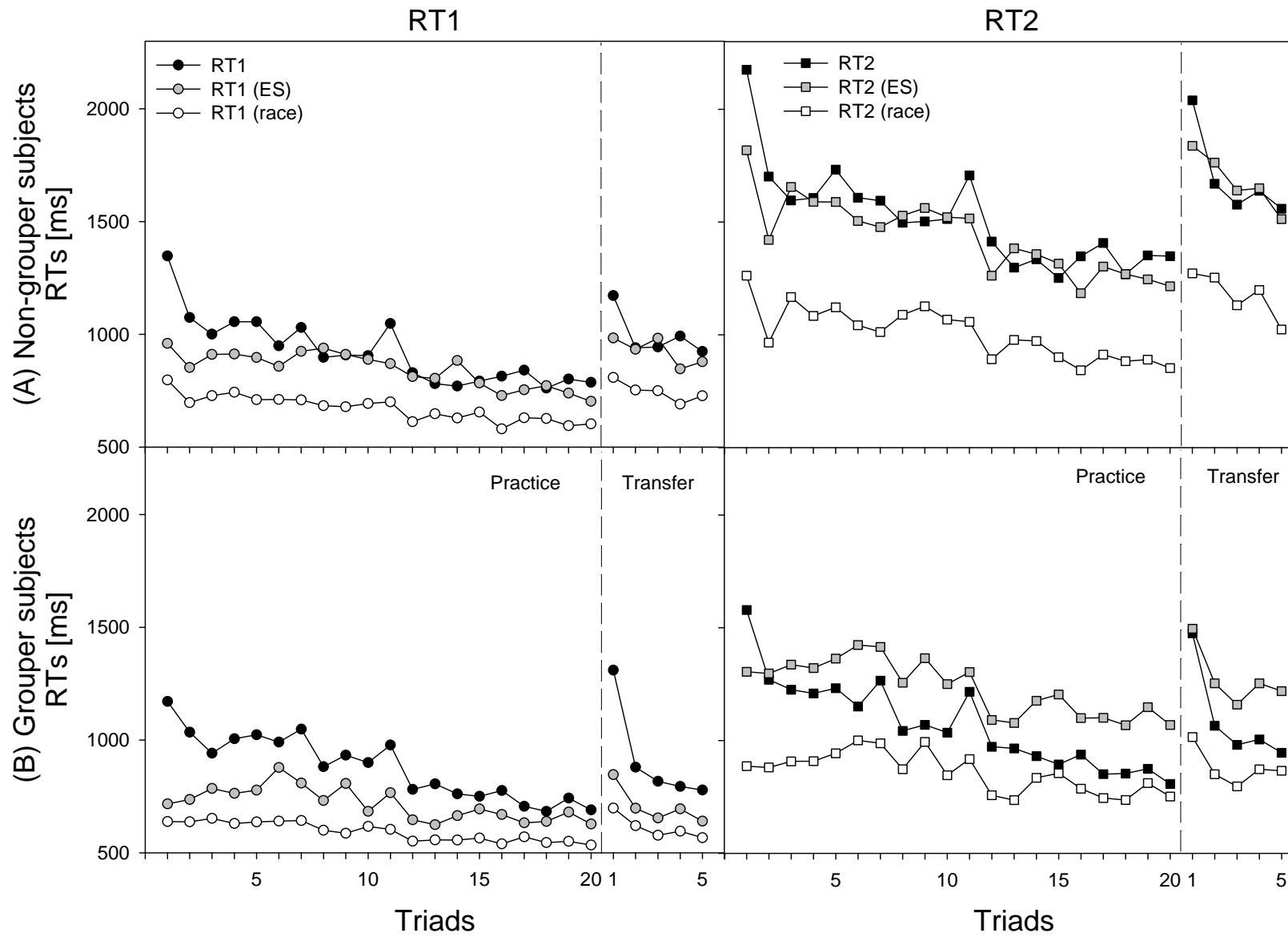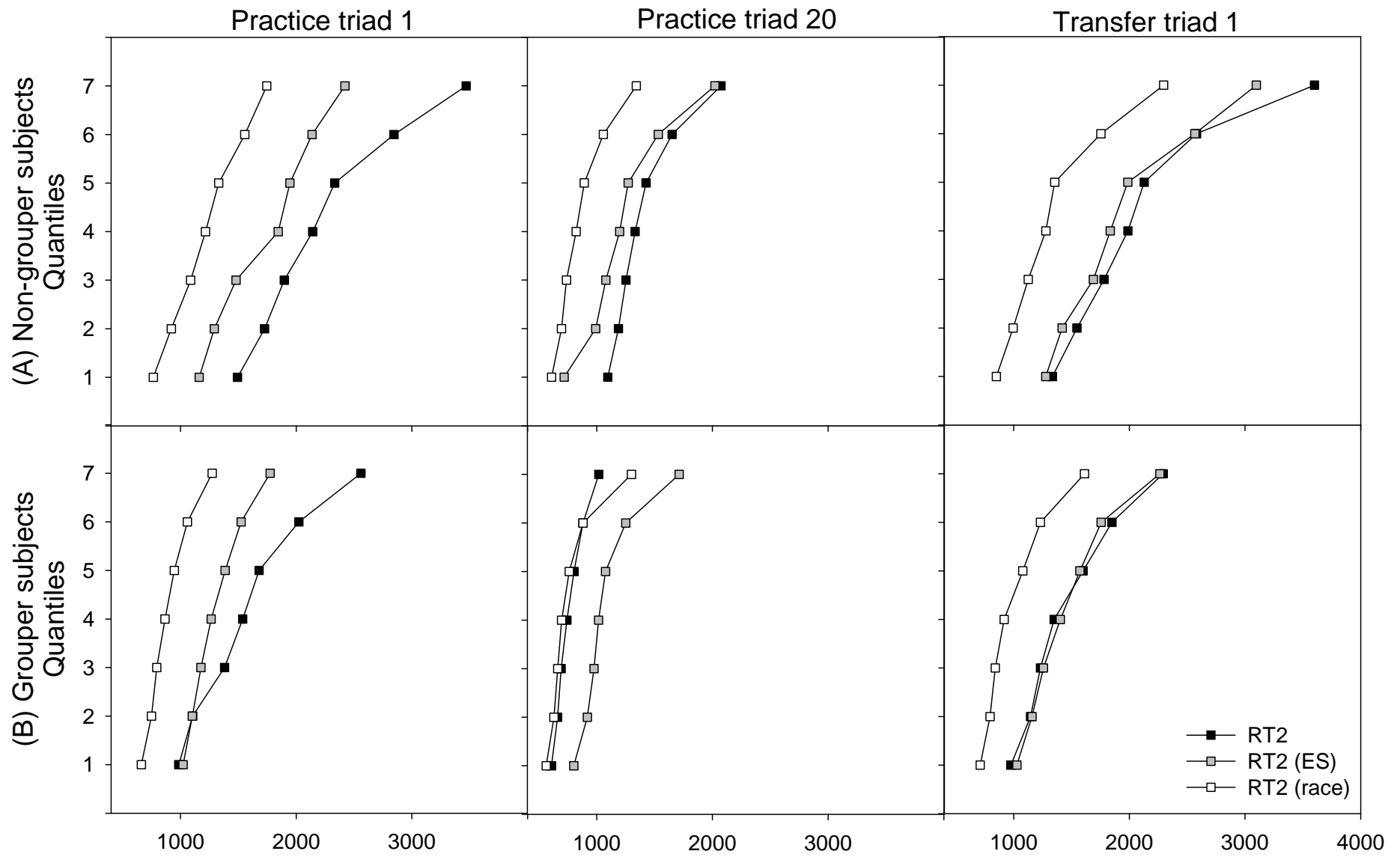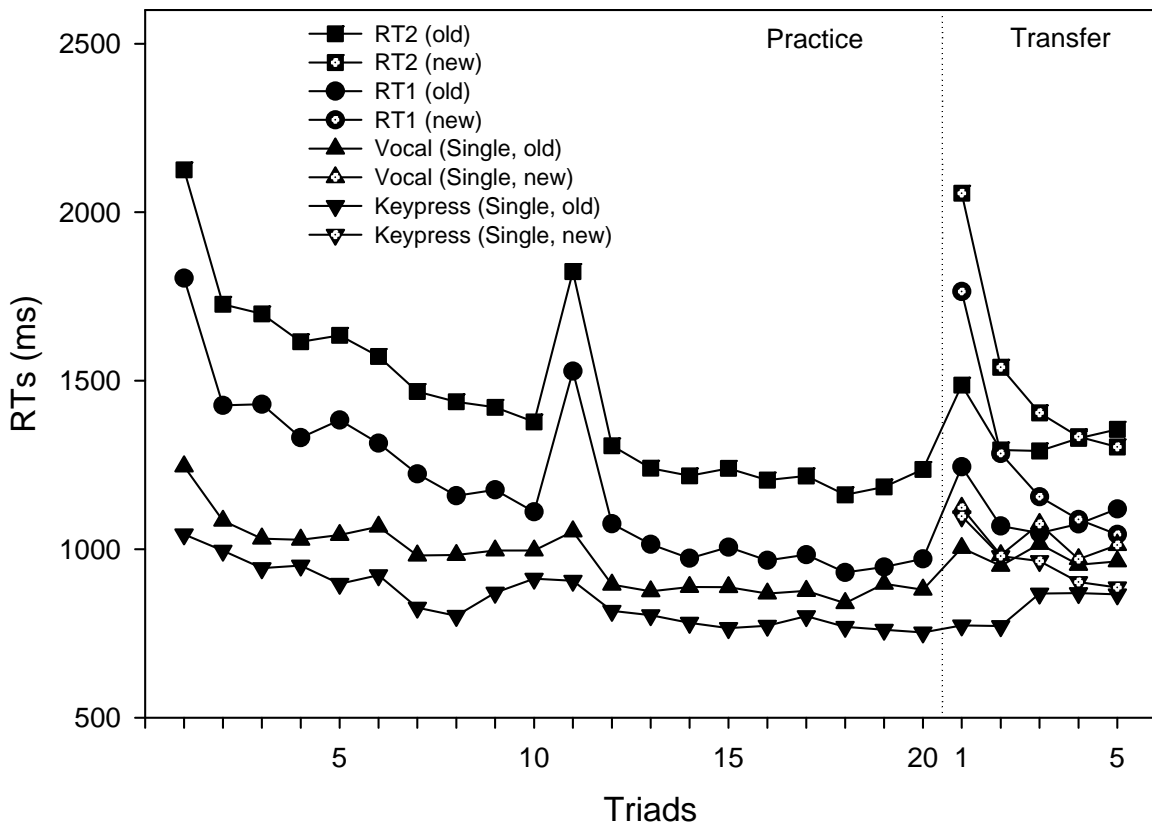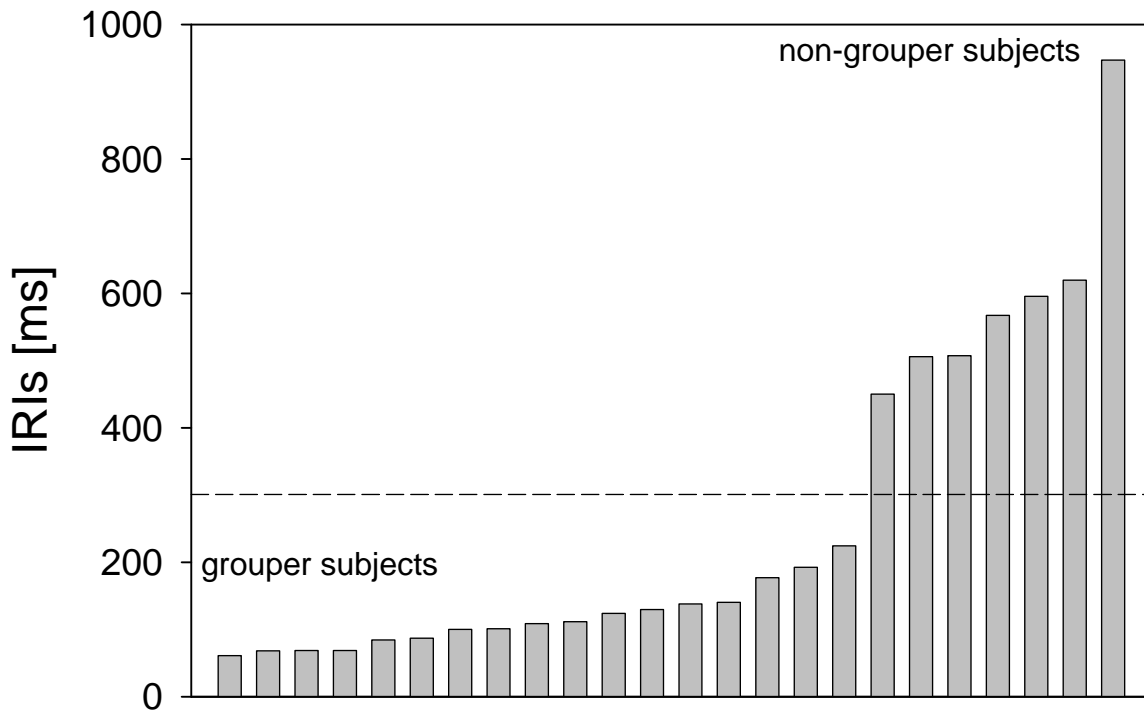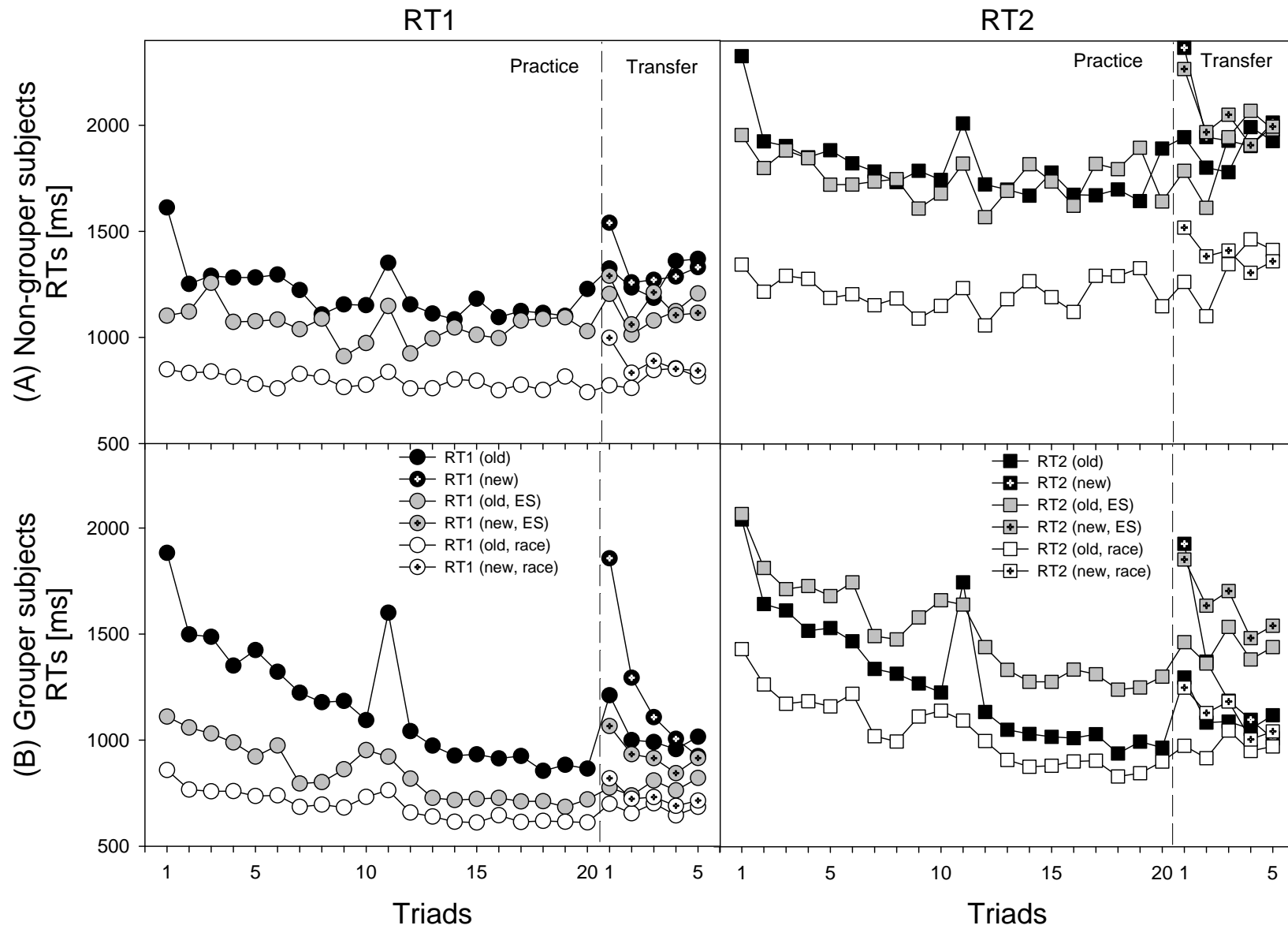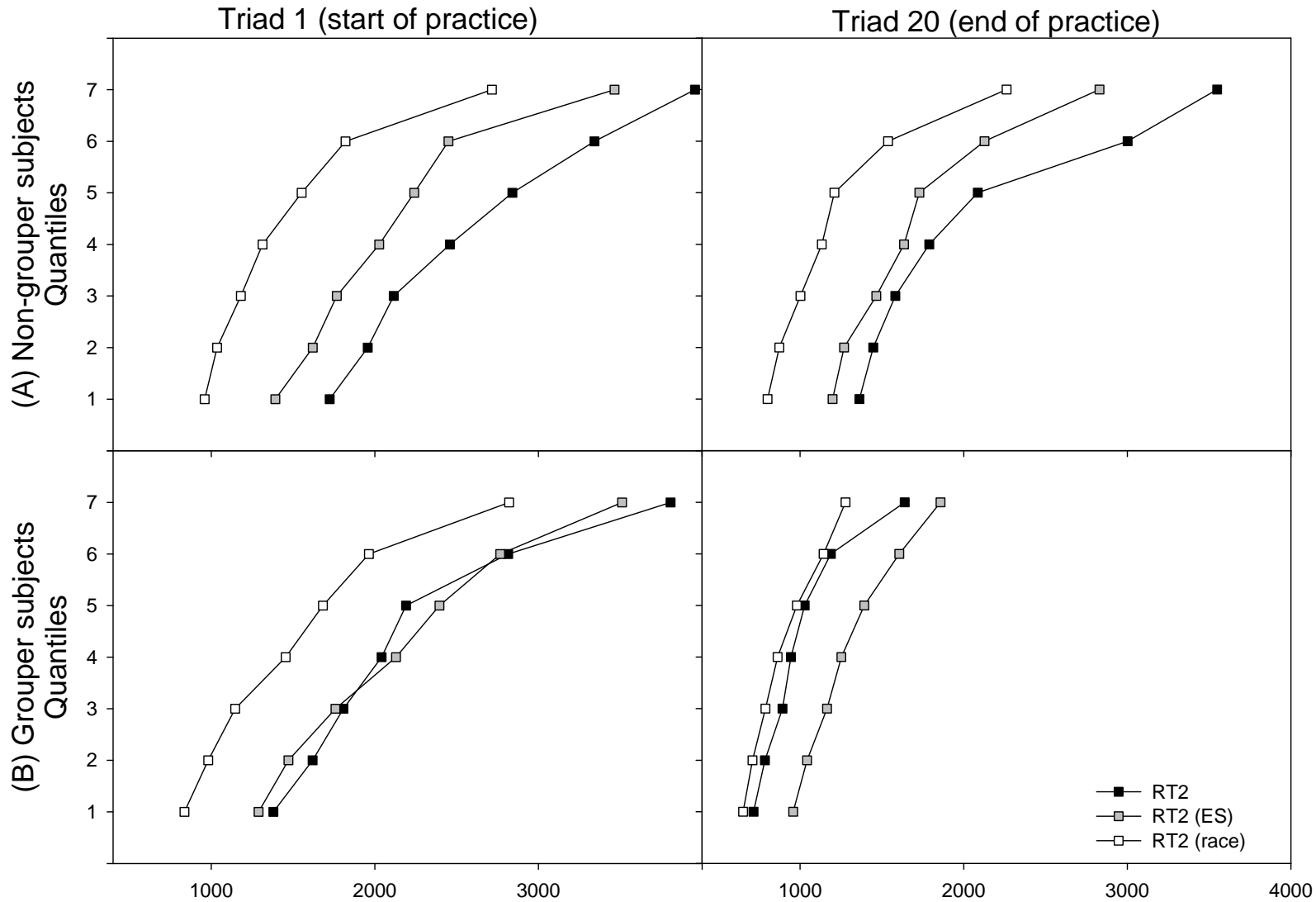
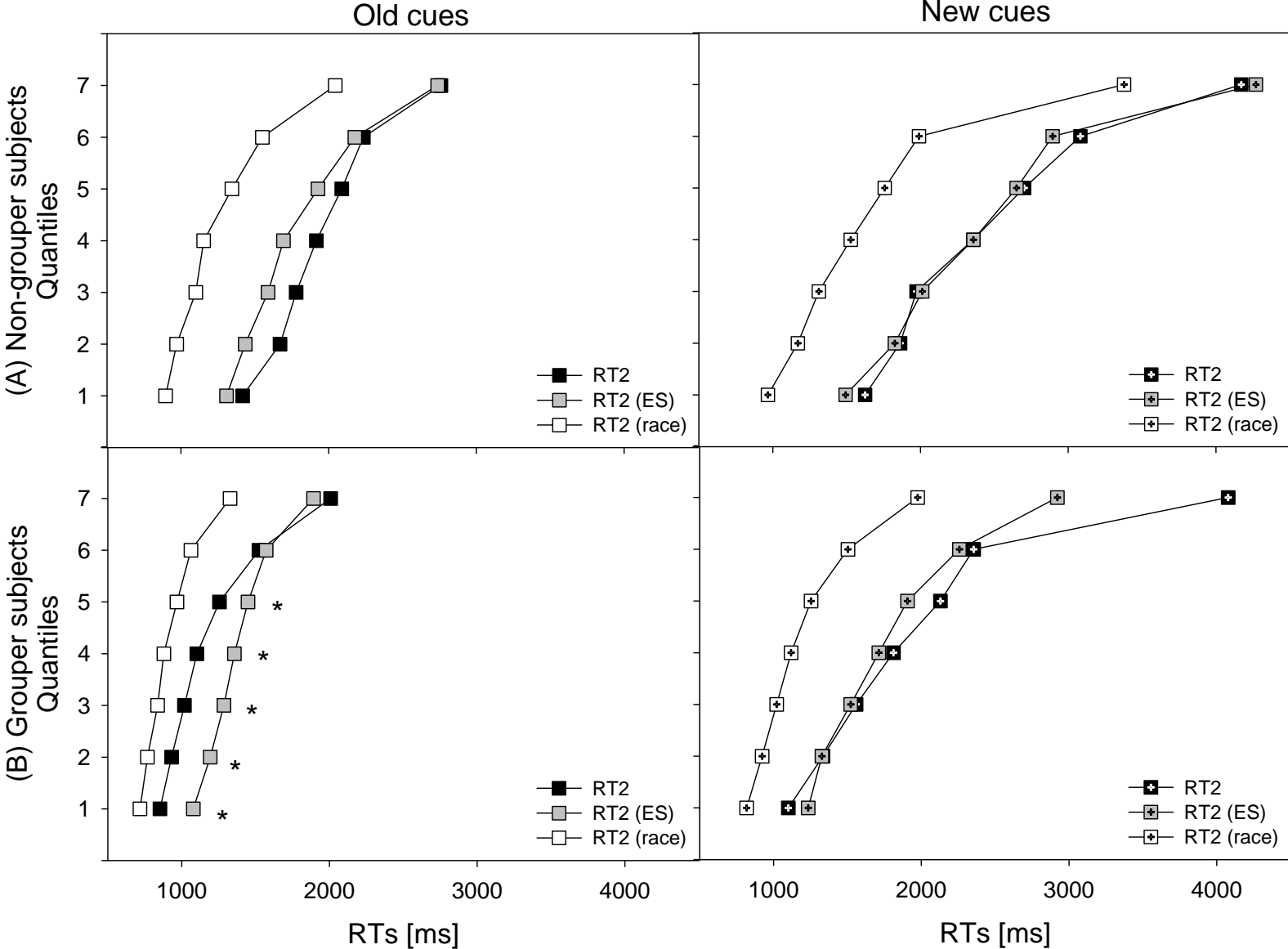Figure 1.
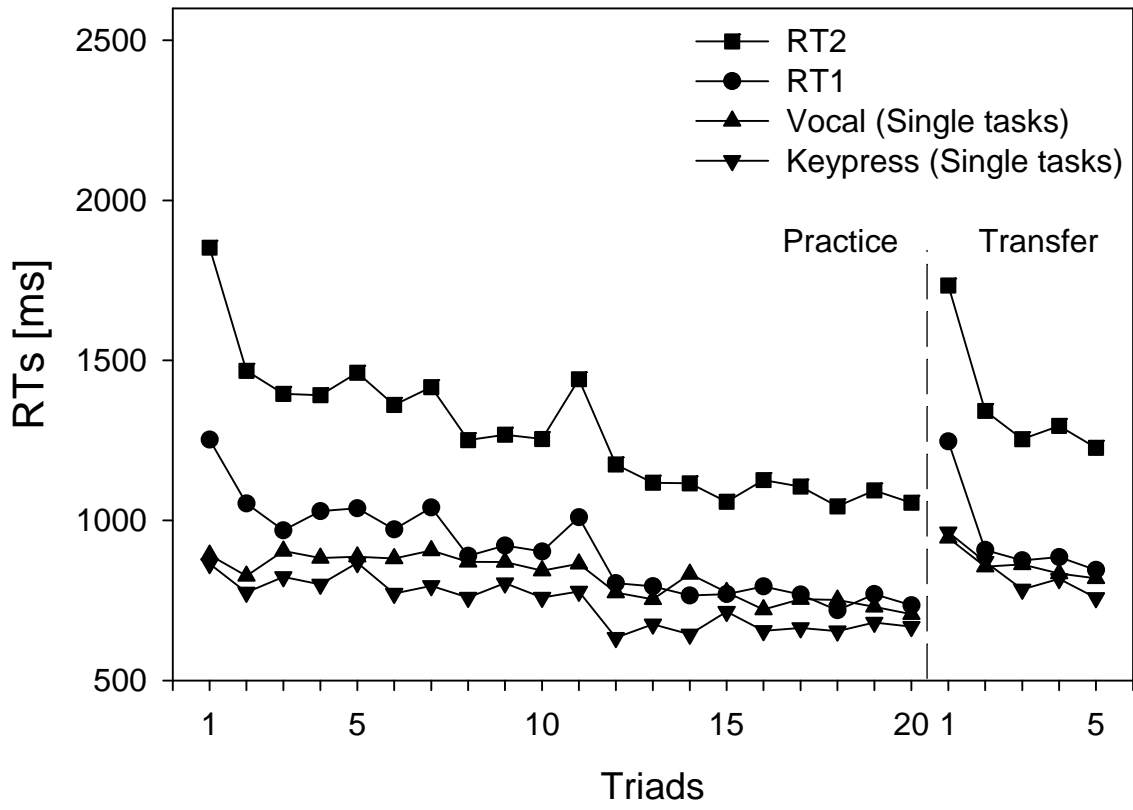
Figure 2.
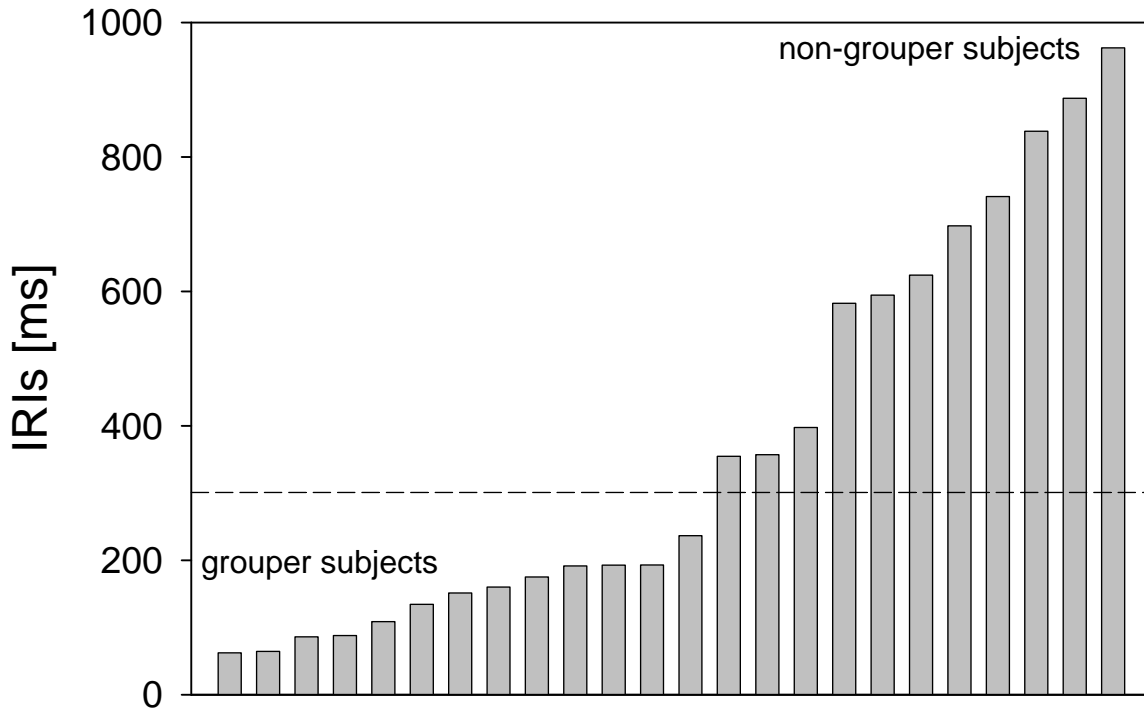
Figure 3.

Figure 4.

Figure 5.

Figure 6.

Figure 7.

Figure 8.

Figure 9.

Figure 10.