Original Article

QJEP

Quarterly Journal of Experimental Psychology
2019, Vol. 72(5) 1005–1028

DOI: 10.1177/1747021818776818
qjep.sagepub.com
SAGE

Two retrievals from a single cue: A bottleneck persists across episodic and semantic memory

Franziska Orscheschek[1], Tilo Strobach[1], Torsten Schubert[2] and Timothy Rickard[3]

Abstract

There is evidence in the literature that two retrievals from long-term memory cannot occur in parallel. To date, however, that work has explored only the case of two retrievals from newly acquired episodic memory. These studies demonstrated a retrieval bottleneck even after dual-retrieval practice. That retrieval bottleneck may be a global property of long-term memory retrieval, or it may apply only to the case of two retrievals from episodic memory. In the current experiments, we explored whether that apparent dual-retrieval bottleneck applies to the case of one retrieval from episodic memory and one retrieval from highly overlearned semantic memory. Across three experiments, subjects learned to retrieve a left or right keypress response form a set of 14 unique word cues (e.g., black—right keypress). In addition, they learned a verbal response which involved retrieving the antonym of the presented cue (e.g., black—"white"). In the dual-retrieval condition, subjects had to retrieve both the keypress response and the antonym word. The results suggest that the retrieval bottleneck is superordinate to specific long-term memory systems and holds across different memory components. In addition, the results support the assumption of a cue-level response chunking account of learned retrieval parallelism.

Keywords

Cued retrieval; dual-retrieval practice; chunked retrieval; bottleneck models

Received: 6 October 2017; revised: 18 April 2018; accepted: 19 April 2018

Introduction

There is an ongoing debate in the fields of cognitive psychology and cognitive neuroscience regarding dual-task performance and its theoretical underpinnings (Fischer & Plessow, 2015). One representative dual-task scenario is the retrieval of two different responses from long-term memory, and a core theoretical question in that case is whether and when such retrieval occurs sequentially vs in parallel. This issue is often accompanied by another question, namely the investigation of a capacity constraint, or bottleneck, that influences the dual-retrieval. Whereas previous research (Strobach, Schubert, Pashler, & Rickard, 2014) has established the presence of a retrieval stage bottleneck for retrieval of two responses from episodic memory even after practice, it remains an open question whether that retrieval stage bottleneck is a global property of long-term memory retrieval or rather only a property of retrieval of two responses from a single memory system; namely, the episodic memory system. That question is at the core of the current research.

Practice effects and dual-retrieval from a single cue: the current state of the art

The bottleneck aspect in dual task research is often explored through the manipulation of different levels of practice. For instance, whereas two episodic memory retrievals have been observed to occur in a sequential manner at the onset of practice, following dual-retrieval practice a subset of subjects exhibit a response time (RT) pattern (at the level of both means and cumulative

[1]Department of Psychology, Medical School Hamburg, Hamburg, Germany
[2]Department of Psychology, Martin Luther University Halle-Wittenberg, Halle, Germany
[3]Department of Psychology, University of California, San Diego, La Jolla, CA, USA

Corresponding author:
Franziska Orscheschek, Department of Psychology, Medical School Hamburg, Am Kaiserkai 1, 20457 Hamburg, Germany.
Email: Franziska.orscheschek@medicalschool-hamburg.de

distributions) that violates an empirically derived lower bound for sequential retrieval, and is instead consistent with learned parallelism at the retrieval stage (i.e., post-perceptual and pre-motor) of processing (Nino & Rickard, 2003; Strobach et al., 2014). For another subset of subjects, however, that transition to parallel retrieval does not appear to occur.

In those experiments, subjects had to associate each of a set of colour-word cues with a unique keypress and a unique vocal-digit response (Strobach et al., 2014). During the keypress learning phase, subjects learned to press a left- or a right-side 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 subject learned to press the left response key and they learned to respond vocally with the word "*five*." After practice to a high level of accuracy of both the keypress and vocal single-retrieval tasks, subjects were given an additional retrieval condition. In this condition, the dual-retrieval task, subjects had to execute both the keypress and the vocal-digit responses that had been associated with the presented colour-word cue. Each of the single- and dual-retrieval conditions were classified as a *block*, which resulted in three different block types: single-retrieval keypress block, single-retrieval vocal block, and dual-retrieval block. Together, these three blocks formed a *triad*. Each cue presentation in each of the blocks in a triad was classified as a *trial*. The triads were repeatedly practised in a so-called *practice phase*. The fact that one response was vocal and the other was manual on trials of the dual-retrieval task should virtually eliminate motor stage response interference (Strobach et al., 2014). In addition, the fact that a single cue was presented on dual-retrieval trials eliminates any interference that may occur in a more traditional dual-task paradigm due to the presentation of a separate cue for each response (Fagot & Pashler, 1992; Strobach et al., 2014). Furthermore, any observed task execution bottleneck or interference is unlikely to reflect working memory demands, because those demands appear to be minimal for this task relative to expected capacity.

Results of these experiments revealed two apparently distinct categories of subjects: *Non-grouper subjects* and *grouper subjects* (Strobach et al., 2014). These subject categories can be assessed by differences in their inter-response intervals (IRIs) on dual-retrieval trials. An IRI refers to the difference between *RT*1 (the latency between cue presentation and the first executed response) and *RT*2 (latency between cue presentation and the second executed response). *Non-grouper subjects* displayed large IRIs on dual-retrieval trials, and hence appear to have executed each response as soon as it was retrieved, whereas *grouper subjects* exhibited small IRIs, and thus appear to have waited until both responses were retrieved, and then synchronised, or "grouped," their response execution.

Over extensive practice, both RT1 and RT2 for non-grouper subjects converged on the quantitative predictions of a sequential retrieval model that we refer to as the *efficient sequential* (*ES*) *retrieval model*. It incorporates the following three assumptions: (1) there are independent and sequential perceptual, retrieval, and motor stages of processing, (2) a bottleneck exists exclusively during the memory retrieval stage of processing, and (3) coordination of the three processing stages during dual-retrieval is maximally efficient (i.e., has no or negligible coordination or task switch delays); see the Appendix A of Strobach et al. (2014) for a mathematical implementation of the ES model. The ES model constitutes an approximate lower bound RT estimate for sequential retrieval stage processing. If that lower bound is systematically and substantially violated by the data, then the sequential retrieval stage hypothesis can be considered false. That lower bound was not crossed by non-grouper subjects in our prior work even after up to 20 blocks of dual-retrieval practice.

In contrast to the *non-grouper subjects*, the response *grouper subjects* were characterised by a small mean IRI on dual-retrieval trials. RT2 for grouper subjects was initially longer than the ES retrieval prediction, suggesting that those subjects retrieved the two responses sequentially and inefficiently at the outset of dual-retrieval practice (just as non-grouper subjects did). In contrast, the mean RT2 fell several hundred milliseconds below the ES prediction by the end of practice, clearly implying learned retrieval parallelism for that subset of subjects. After extended practice, RT2s for those subjects approached, but remained slightly above, a simple race model of parallel processing; see Appendix A of Strobach et al. (2014) for the implementation of this model. In sum, it can be concluded that some type of *learned retrieval parallelism* occurred for the response grouper subjects. However, that result appears to depend on two conditions jointly holding (1) response grouping and (2) dual-retrieval practice.

The results described above raise the question of what facilitated the onset of parallel retrieval for grouper subjects, and in particular whether it is a cue-level or a task-level phenomenon (Strobach et al., 2014). One possibility is that response grouping subjects underwent a strategic and global switch from sequential to parallel retrieval following modest dual-retrieval practice (Meyer & Kieras, 1997; Oberauer & Bialkova, 2011; Oberauer & Kliegl, 2004). According to the alternative *set-cue bottleneck model* (Figure 1), however, grouper subjects learned to chunk the two responses independently for each cue, such that both responses could be retrieved in one pass through the same retrieval bottleneck that appears to govern performance at the beginning of practice, and throughout practice for non-grouper subjects (Nino & Rickard, 2003). That chunking process may only be possible when both responses are concurrently in working memory, as should be the case for response grouper subjects within the set-cue bottleneck

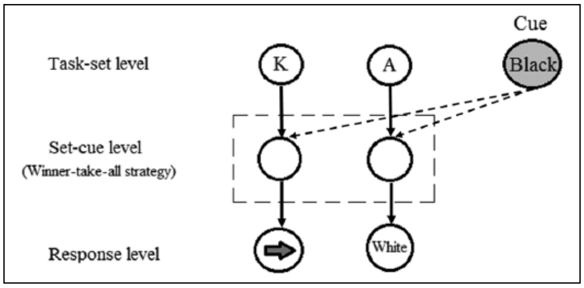

Figure 1. Associated processing levels of the set-cue bottleneck model. The model presumes distinct depictions for the cue and the task-set at the task-set level (i.e., either the keypress response [K] or the antonym response [A]). During learning, the set-cue level emerges which represents the connection of each cue and the associated response. This results in the connected response level for each cue-response pairing. The set-cue level incorporates the bottleneck, since only one node at the set-cue level can influence performance at time in a dual-retrieval condition. Individuals thus have to complete both responses sequentially.

framework but not, or much less so, for non-grouper subjects. By using the chunking process, subjects are able to form just one mental representation that can be in the main focus of attention during dual retrieval. Therefore, groupers do not need to divide their attention between two foci because the potential bottleneck-like foci and the switch between them (Cowan, 2010; Garavan, 1998; Oberauer, 2002) play no role in the processing of chunked retrievals. Furthermore, an advantage of the bottleneck model is that it provides a natural account, rooted in associate and working memory theory, of why learned parallelism appears to be restricted to the case of response grouping (Nino & Rickard, 2003; Strobach et al., 2014).

Because response chunking is an associative memory process that would occur during a dual-retrieval trial, the response chunking account suggests that learned parallelism occurs at the individual cue-response level rather than at the task level. Consistent with that possibility, it was shown (Strobach et al., 2014) that learned parallelism for grouper subjects did not transfer to cues for which only single-retrieval practice occurred during the practice phase, with dual-retrieval trials introduced only during a transfer phase. This transfer phase had the same structure as the practice phase (i.e., triads of two single and one dual-retrieval blocks). In order to assess the retrieval process, dual-retrieval trials in the transfer phase included cues that were previously only practised in single-retrieval trials (i.e., new cues; Strobach et al., 2014). The finding that learned parallelism did not transfer to these new cues is analogous to expert memory of expert chess players, who tend to chunk their knowledge about single components of complex chess trains (Chase & Simon, 1973). Whereas expert chess players are able to recall complete sets of well-known and practised chess trains by this method of chunking, they are not able to perform this complete set recall on novel chess trains that they did not practice before (Gong, Ericsson, & Moxley, 2015). In general, these lines of research support the cue-level response chunking of learned retrieval parallelism over the strategic task-level account. Strobach et al. (2014) suggested that the set-cue bottleneck model holds in all cases described above (i.e., for non-grouping and grouping of responses), and that cue-level response chunking with practice allows both retrievals to occur in one pass through that bottleneck.

A dual-retrieval bottleneck with learned parallelism through chunking: a generalisable phenomenon?

Since the previous studies by Nino and Rickard (2003) as well as Strobach et al. (2014) used novel cue-response associations, it remains an open question if the observed dual-retrieval bottleneck would also be present in retrieval situations that involve automatised cue-response associations. In order to assess dual-memory retrieval mechanisms that involve an automatised task, we investigated these mechanisms with a new task set. This task-set comprised novel cue-response associations as well as well-established, and automatised, associations. In detail, the new dual-retrieval task includes a cue word that is one member of an antonym pair (e.g., the word *black*) and the two required responses are (1) the other member of the antonym pair (e.g., vocalisation of "white") and (2) a newly acquired, episodic left or right keypress response associated with the same retrieval cue (i.e., the same keypress task as in our prior work, but cued by a member of an antonym pair, i.e., *black*, rather than a colour word cue).

The rationale for using antonym pairs is that the verbal antonym retrieval might differ substantially in the level of automatisation and therefore also in the level of processing in comparison to the keypress retrieval. Antonyms can be classified as strongly semantically related words that are known to cue each other automatically in semantic priming tasks (Lucas, 2000; Perea & Rosa, 2002). In a typical semantic priming task, subjects are required to make lexical judgements to target words (i.e., "Is the following item a `word` or a `nonword`?") (Hutchison, 2003). The associated *sematic priming effect* indicates that subjects react more accurately and rapidly to a target word when a semantically and associatively related word was presented previously (e.g., an antonym word) than when the presented word was semantically unrelated (Hutchison, 2003; Perea & Rosa, 2002). Previous research established that priming occurs automatically for associatively connected word pairs; and that antonyms with a strong forward associative strength (FAS) will be cued in an automatic way (Hermann, Conti, Peters, Robbins & Chaffin, 1979; Lucas, 2000; Perea & Rosa, 2002). FAS refers to the probability that a specific cue word yields a specific response during a free association task (Nelson, Dyrdal, & Goodmon, 2005). Since antonyms can differ in their degree of antonymy (Hermann et al., 1979), we used a group of antonym pairs with strong FASs. It could be argued, that the keypress task also transforms into an automatically cued cue-response association after extensive task practice. However, a microanalytic review by Hutchison (2003) concluded that newly acquired item pairs do not indicate automatic priming except for situations in which they are able to build a meaningful association. Furthermore, such an association is difficult to obtain when the item words are already linked to a set of other words or reactions, like in the case of the keypress task in combination with the antonym pairs (Hutchison, 2003). The same applies to the task sets used in Strobach et al. (2014) as well as in Nino and Rickard (2003): Since both tasks involved newly acquired item pairs and each cue word was paired with one vocal-digit and one keypress response, there should have been no chance for the development of automatic associations.

Hutchison (2003) further discussed, that automatised and well-established links between words and concepts are rooted in the semantic memory system, whereas newly acquired information is stored in the episodic memory system. Studies that used semantic primes found evidence for automatic priming, whereas episodic memory primes did not show such effects (Hutchison, 2003). This leads us to another potential dimension of this new task set-up: Since an important variable in memory research is the type of memory that needs to be retrieved, our new task combination might allow us to assess dual memory retrieval in two different memory domains. These distinct types of memory are the above mentioned episodic and semantic systems (Tulving, 2002; Tulving & Hastie, 1972). Both keypress and vocal response learning in the aforementioned experiments by Strobach et al. (2014) constitute new learning in episodic memory. In the present study, the keypress task would still reflect a retrieval task that could involve episodic memory components. Since semantic memories include decontextualised information about language and associated perceptual attributes, knowledge about antonyms could be classified as semantic memories (Loureiro & Lefebvre, 2016) and the antonym retrieval task could be a retrieval task that is largely semantic. Even though we refer to the antonym retrieval task as a semantic memory task, we want to be careful with the claim that this task is of merely semantic nature. Since studies from different research domains have shown that there could be an increase in episodic memory retrieval components for semantically related materials due to extensive training of a specific memory retrieval (Reder & Ritter, 1992), there might be a chance for episodic memory component involvement in the verbal antonym task after practice.

There are two plausible outcome possibilities of these experiments. On the one hand, dual-retrieval from automatised antonym cues may occur in parallel with no retrieval stage bottleneck, even prior to dual-retrieval practice. If this is the case, we would conclude that the dual-retrieval bottleneck is not global, but may be rather specific to dual-retrieval of novel cue-response associations and thus episodic memory retrieval combinations. On the other hand, we might find results that reflect the processing patterns of previous studies by Strobach et al. (2014). In this scenario, we could observe sequential processing at the onset of dual-retrieval practice and a potential shift to learned parallelism after practice for both task-sets (i.e., verbal antonym retrieval task, keypress task).

One additional reason why a retrieval bottleneck may not be observed in the current experiments is that the antonym retrieval task is highly practised and automatised. Prior evidence from dual choice RT tasks (Maquestiaux, Laguë-Beauvais, Ruthruff, & Bherer, 2008; Ruthruff, Van Selst, Johnston, & Remington, 2006) explored the controversial question of whether the central processing bottleneck in that domain can be bypassed through task automatisation. The authors assessed dual-task performance using the psychological refractory period (PRP) paradigm, in which a highly practised auditory-vocal task (high automatisation condition) was presented along with an unpractised visual-manual task (low automatisation condition). The results supported the assumption of a bottleneck bypass in the high task automatisation condition: Most of the subjects (17 out of 20) exhibited parallel choice processing. In contrast, in the lower task automatisation condition such evidence was only present for very few subjects (Ruthruff et al., 2006).

Considering the possibility that both tasks could reflect two different memory domains, the *multiple memory systems* framework (Tulving & Hastie, 1972) might be seen as consistent with the possibility that there could be no retrieval bottleneck in the current experiments. That framework incorporates the idea that episodic and semantic memories are two distinct memory systems that are specialised in the processing of differential information (Rajah & McIntosh, 2005). It is assumed that both systems involve distinct anatomical and functional substrates of the brain. This concept has been mainly supported by clinical trials which investigated the behaviour and neural processing in patients with memory impairments (e.g., semantic dementia, bilateral hippocampal damage, autism spectrum disorder, etc.) (Gaigg, Bowler, & Gardiner, 2014; Rajah & McIntosh, 2005; Schmolck, Kensinger, Corkin, & Squire, 2002). One example from these studies is the presence of abnormal functioning of episodic memory in autism spectrum disorder, but normal functioning of semantic memory (Gaigg et al., 2014). Another example is that patients with bilateral hippocampal damage sustained in adulthood do not exhibit deficits in semantic memory, such as antonym retrieval, but have difficulties with episodic memories (Schmolck et al., 2002). According to these indications of the *multiple memory systems* account, when the two required retrievals are potentially rooted in two different memory domains, it could be assumed that there would be no retrieval bottleneck (although that theory does not explicitly make that prediction).

In addition, the scarce research on dual-retrieval with semantic memories found evidence in favour for the presence of parallelism and against the presence of a sequential bottleneck processing. In a study that used the PRP paradigm, Logan and Schulkind (2000) used a combination of a letter and a digit task to assess semantic memory retrieval across four different experiments. They assumed that if parallel retrieval was present, they should find that information retrieved from the second stimulus would influence processing of the first stimulus, i.e., crosstalk between the two tasks. Their results supported the presence of crosstalk for semantic memory.

In contrast to the multiple memory systems account, however, the *unitary memory system* account suggests that the retrieval bottleneck might hold in the current experiments. The unitary memory system account states that the constructs of episodic and semantic memory are endpoints of a single memory "spectrum" that is manifested in a unitary brain system (García-Lázaro, Ramirez-Carmona, 2012; Rajah & McIntosh, 2005). These distinct ends of the spectrum are assumed to account for overlapping as well as differential retrieval processing in an array of neural substrates (García-Lázaro et al., 2012). Rajah and McIntosh (2005) supported this perspective with a neuroimaging experiment that indicated similar interactions of brain systems that are independent of the type of memory retrieved. These results support the view of a single network with overlapping processes. Applying these results to our new experiments could lead us to the suggestion that, if there is a *unitary memory system* for episodic as well as semantic memories, the outcome of a dual-retrieval task that combines episodic and potentially semantic retrieval might be the same as for a task that involved two episodic memory retrievals, i.e., that a retrieval bottleneck might be observed in the current experiments both prior to and after practice.

Summarising these findings, high levels of automatisation seem to favour parallel rather than bottleneck-based (or other limited capacity) models (Meyer & Kieras, 1997). It is possible that similar automaticity effects hold for dual long-term memory retrieval when one of the tasks is highly automatised, as in the current case of antonym retrieval (Lucas, 2000; Perea & Rosa, 2002). With the use of the new task set-up, we can enable a more comprehensive examination of the nature and generalizability of a dual-retrieval bottleneck since we might find differential outcomes with an automatised versus a novel cue-response association than with two novel associations.

Experiment 1

The general procedure of Experiment 1 is similar to Experiment 1 of Strobach et al. (2014): After dual-retrieval practice on one half of the retrieval cues (i.e., old cues), we tested dual-retrieval performance on the alternative cues (i.e., new cues) during transfer in a context without old cues. If the combination of novel and automatised cue-response associations is subject to the same bottleneck as is dual-retrieval of two novel cue-response combinations, then results of this experiment should mirror those of Strobach et al.: (1) dual-retrieval RTs should be above the ES model predictions at the outset of practice, (2) an onset of parallel retrieval should be observed with practice, and (3) there should be no transfer of learned parallelism to newly presented dual-retrieval cues on the transfer tests. Alternatively, the retrieval bottleneck model may not hold for the case of mixed novel and automatised retrievals. In that case, parallel retrieval may be evident as early as the first dual practice trial, and it should also be evident for new dual-retrieval cues on the transfer test.

Methods

Subjects. The experiment included 24 undergraduate students of the University of California, San Diego. The sample consisted of 17 female and 7 male subjects who had a mean age of 21.3 years (SD = 2.8). All of the subjects had normal or corrected to normal vision and were right handed. They received course credit for their participation and were naïve of the study's research aim.

Apparatus and cues. 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 antonym responses and manual keypress responses were recorded with the accompanying voice-key apparatus (Model 200A). The voice-key apparatus included a serial response (SR) box that incorporates a debounce period of 0 milliseconds for a keypress reaction (Psychology Software Tools, Pittsburgh, PA). The SR box measures and hallmarks keypress RT data and assess the reaction for its accuracy. Verbal reactions were measured with the accompanying microphone that was connected to the SR box. The list of cues, the corresponding keypress responses in the keypress task, and the antonym responses in the antonym task, are shown in Table 1. The cues were presented on a 19" CRT monitor with an estimated visual angel of 51.5° from a viewing distance of 50 cm. Cue words subtended to 7 cm and letter height was 1.7 cm which relates to a visual angle of 8.0° and 1.9°, respectively.

As mentioned in the introduction, antonyms can differ in their degree of antonymy (Hermann et al., 1979). This refers to the degree of their relationship and their FAS. Since antonym pairs with high levels of antonymy evoke

Table 1. Cue-response pairs: antonym cue words (English [Experiment 1]/German [Experiments 2 and 3]), antonym response words (English [Experiment 1]/German [Experiments 2 and 3]), and keypress responses.

Antonym cues	Antonym responses	Keypress response
Bad/Schlecht	Good/Gut	←
Black/Schwarz	White/Weiss	→
Day/Tag	Night/Nacht	→
Dry/Trocken	Wet/Nass	→
East/Ost	West/West	←
Fast/Schnell	Slow/Langsam	→
Hate/Hass	Love/Liebe	←
Hot/Warm	Cold/Kalt	←
In/Innen	Out/Außen	→
Lost/Verloren	Found/Gefunden	→
Low/Niedrig	High/Hoch	←
Rich/Reich	Poor/Arm	←
Thick/Dick	Thin/Dünn	→
True/Wahr	False/Falsch	←

quicker responses (and their probability of automatisation might be increased) than antonym pairs with low levels of antonymy (Hermann et al., 1979; Nelson et al., 2005), we wanted to ensure a high degree of antonymy. Hutchison et al. (2013) set up an extensive database (http://spp.montana.edu) in which 1661 target words and primes were assessed in a lexical decision task as well as a speeded naming task. As we wanted to ensure robust FAS levels, we checked for our chosen antonym pairs in the database. We found 11 out of our 14 antonym pairs in the list. All of these antonym pairs had moderately high to very high FAS values (μ = 0.67, *SD* = 0.1). This mean value reflects an overall high level of FAS for our antonym pairs (Nelson et al., 2005). We further looked at the values that these 11 pairs obtained in a latent semantic analysis (LSA, Landauer & Dumais, 1997; Hutchison et al., 2013). LSA refers to an analysis of the relationship of a set of documents and containing terms. It is assumed that related terms would appear more often in documents dealing with the same or related topics (i.e., their similarity is based on their co-occurrence) (Simmons & Estes, 2006). Specifically, LSA analyses the universal co-occurrence of a prime and a target word (Hutchison et al., 2003; Landauer & Dumais, 1997). Similarity estimates based on LSA in the database of Hutchison et al. (2003) showed that our 11 antonym pairs reflected a mean LSA value of 0.68 (*SD* = 0.13). Since LSA values in the range of 0.67 to 1.00 are considered to indicate a strong association (Simmons & Estes, 2006), our antonym pairs can be classified as having a moderately strong association in the mean.

Moreover, since we wanted to establish a task-set that differentiates significantly from the task combination in Strobach et al. (2014), we conducted an assessment of the conceptual connection between the items in the antonym retrieval task and the items in the vocal retrieval task in Strobach et al. Conceptual connection referred to the level of association between the cues and responses. Hence, we presented a total of 86 first year psychology students (61 female and 24 male; mean age 21.8 years, *SD* = 2.91) with a list of the 14 antonym pairs used in the present experiments, and the 14 colour-digit combinations used in Strobach et al. Subjects were required to assess the level of conceptual connection on a scale from 1 (i.e., "0% association between cue and response") to 7 (i.e., "100% association between cue and response"). A paired samples t-test showed a significant difference in the ratings of the association levels between both sets of cue-response combinations, $t(86) = 52.71$, $p < .001$. Whereas the mean score of association for the antonym pairs was 6.45 (*SD* = 0.65) which reflects a relatively high level of conceptual connection, the mean score for the colour digit pairs was 1.39 (*SD* = 0.58) which reflects a rather low level of association. These results clearly demonstrate the distinction between both sets of cue-response combinations and highlight our claim that the antonym pair combinations are well established.

Procedure and design. An overview of the design is given in Table 2 and Figure 2. Session 1 started with a *study phase.* During this study phase, subjects were introduced to the keypress task, in which they were instructed to memorise the 14 cues and the associated manual responses (see Table 1). The associated direction for each cue word was indicated with an arrow on the screen that pointed in either the right or left direction on the display. The subjects were instructed to memorise each cue response pair and to press an associated key on the keypress pad. Every cue-response combination was presented once, randomly ordered, in each of 2 study blocks. Each cue presentation is classified as a *trial*. The specific trial procedure was as following: on each trial, the cue and the corresponding arrow 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 arrow and subjects were instructed to press the associated key.

After these two study blocks, the subsequent *single-retrieval criterion phase* started. Each cue word was again presented once, without the directional response cue, and subjects were instructed to retrieve the earlier associated left of right 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. If the response was correct, the next trial began immediately thereafter. If the response was incorrect, an "incorrect" message, plus the correct response (in form of a left or right oriented arrow), was presented for 2,500 ms. The presentation of all 14 cues (i.e., 14 trials) is classified as a *block*. This phase included 15 of these single-keypress blocks.

Table 2. Overview of the general procedure.

	Experiment 1/2		Experiment 3	
Session	Experimental phases	Number of blocks/triads	Experimental phases	Number of blocks/triads
1	Study phase (keypress task)		Study phase (keypress task)	
	Single-retrieval criterion phase (keypress task)	15 blocks	Single-retrieval criterion phase (keypress task)	15 blocks
2	Single-retrieval criterion phase (keypress task)	2 blocks	Single-retrieval criterion phase (keypress task)	2 blocks
	Study phase (antonym task)		Study phase (antonym task)	
	Single-retrieval criterion phase (antonym task)	2 blocks	Single-retrieval criterion phase (antonym task)	2 blocks
	Single–dual practice phase	15 triads	Single–dual practice phase	20 triads
	Single–dual transfer phase	5 triads		
3			Single–dual practice phase	20 triads
			Single–dual transfer phase	5 triads

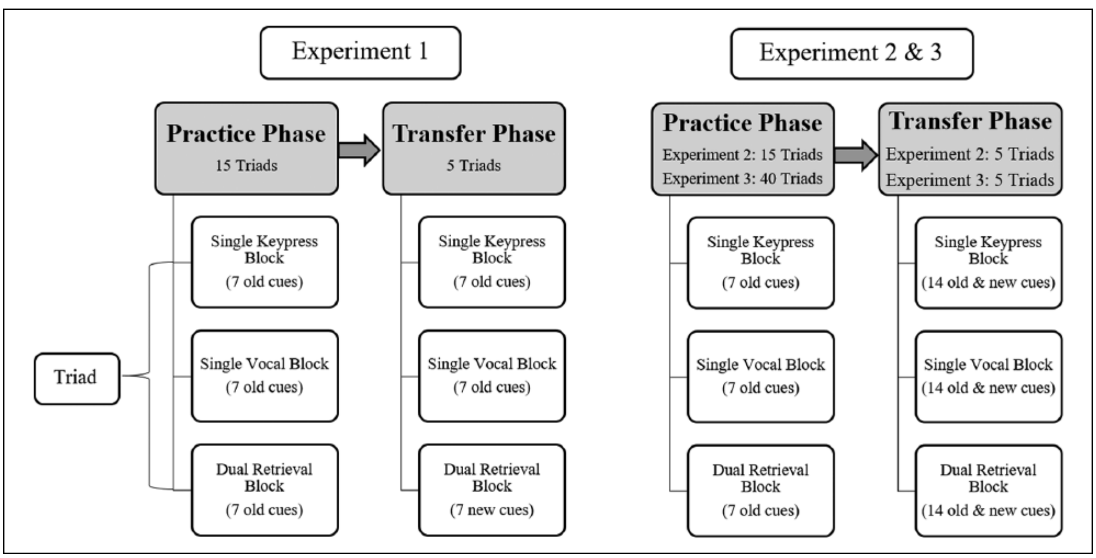

Figure 2. Overview of the experimental designs across all experiments.

In Session 2, two blocks of the single-keypress task were repeated before the *study phase* of the antonym task was started (note that this latter task was not performed in Session 1). The antonym task study phase was identical to the study phase for the keypress task with the exception that the cue word and the associated response word were presented and subjects were instructed to speak that response word clearly into the microphone. The cue response mappings in this task involved the same 14-word cues as the keypress task. However, the associated responses were the antonyms to each of the words (e.g., Black—White; East—West; Hot—Cold, etc., see Table 1). In this task, experimenters coded for response correctness on the keyboard of the experimental computer. Similar to the keypress feedback, in case of an incorrect response, an "incorrect" message accompanied by the correct response were presented on the screen. Therefore, an interval of 2,500 ms was introduced after the execution of the vocal response. The *single-retrieval criterion phase* of the antonym task was identical to that phase in the keypress task with the exception that subjects were instructed to produce the vocal response words. This phase included 2 blocks, with 14 trials in each block.

Following the single-retrieval criterion phase, subjects had to complete the *single-dual practice phase*. This phase consisted of 15 *triads* (Practice Triads 1-15). A triad is a set of 3 blocks which refer to the keypress task, the antonym task, and dual-retrieval. The keypress as well as the antonym tasks are referred to as single-retrieval blocks and the dual-retrieval task is referred to as the dual-retrieval blocks. Within each triad, these blocks were presented in 2 different block orders: One half of the subjects performed the antonym task in the first single-retrieval block, the keypress task in the second single-retrieval block, and dual-retrievals in the

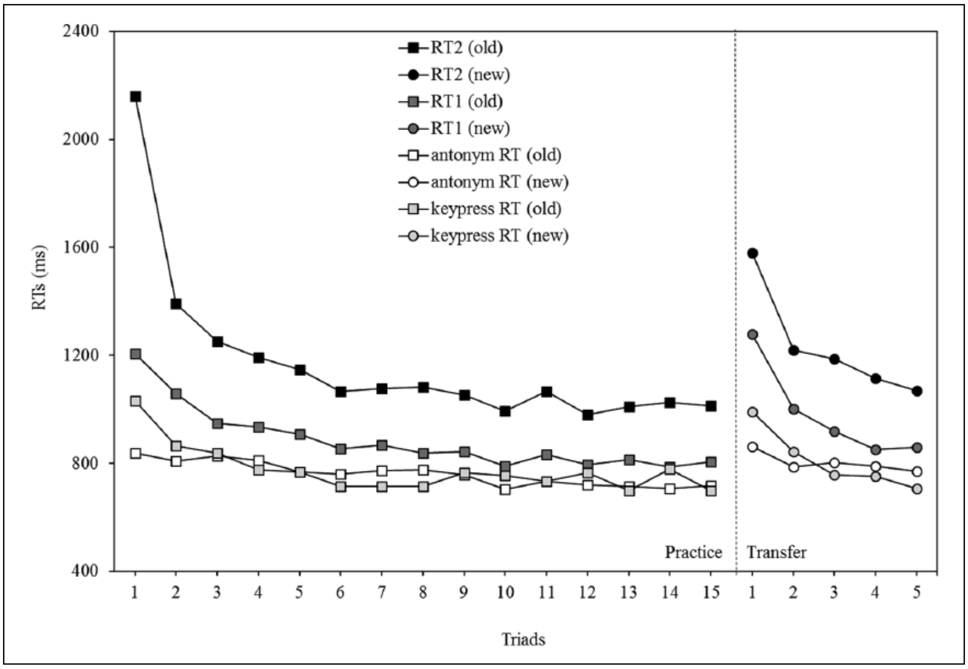

Figure 3. Observed reaction times (RTs) in single-retrieval blocks of the keypress task and the antonym task as well as observed RTs in dual-retrieval blocks (i.e., RT1 and RT2) in the overall dataset during the 15 practice triads and 5 transfer triads in Experiment 1.

third block. The remaining subjects performed the reversed block order with a first dual-retrieval block, a second keypress single-retrieval block, and a third antonym single-retrieval block. The trials in the single-retrieval blocks were identical to those in the single-retrieval criterion phases. The dual-retrieval trials were identical to single-retrieval trials with the following exceptions: Subjects were instructed to speak the response word and press the key as quickly as possible while being accurate. The cue remained on the screen until subjects executed both responses. In each block, subjects were presented with 7 trials, one trial for each of 7 of the previously trained 14 cues. Half of the subjects were presented with the 7 cues at the even positions of the list presented in Table 1, while the remaining subjects were presented with the 7 cues at the odd positions of this list during the single-dual practice phase. Each of the 7 cues was presented once per block in randomised 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 single-dual practice phase, the only exception being that the 7 cues that were presented in the single-dual practice phase were not presented in the transfer phase, whereas the 7 cues that were not presented in the single-dual practice phase were presented in the transfer phase.

Results and discussion

Accuracy results. In all analyses, we excluded trials with RTs below 200 ms in order to ensure that voice key failures were excluded from RT analyses (0.2% of all single- and dual-retrieval trials). Voice key failures refer to situations in which the voice key tripped inappropriately (e.g., the voice key might be falsely activated by an accidental cough). In the single-dual practice phase (i.e., from Practice Triad 1 to Practice Triad 15), error rates decreased from 9.5% to 1.0% for the keypress single-retrieval trials, and from 3.4% to 0.1% for the antonym single-retrieval trials. For the keypress task in the dual-retrieval blocks of this phase, error rates decreased from 13.9% to 1.2%, and for the antonym task from 4.5% to 0.1%. Dual-retrieval error rates during the single-dual practice phase decreased from 9.2% to 0.1% for the first completed response (keypress or antonym), and from 12.1% to 0.1% for the second response. Hence, error rates were similar on single- and dual-retrieval trials at the end of practice, albeit slightly higher on dual-retrieval trials at its start.

RT results. RTs averaged over all subjects for correctly performed single-retrieval (i.e., keypress task, antonym task) and dual-retrieval trials (RT1, RT2) are shown in Figure 3. Both responses had to be correct in order to be classified as a correct dual-retrieval. RTs decreased steadily over the course of single-dual practice (Practice Triad 1-15), but increased markedly on the transfer test, particularly for dual-retrieval trials and Transfer Triad 1.

IRI analysis. Following analyses in Nino and Rickard (2003) as well as Strobach et al. (2014), we computed mean IRIs on dual-retrieval trials for each subject, averaging over all

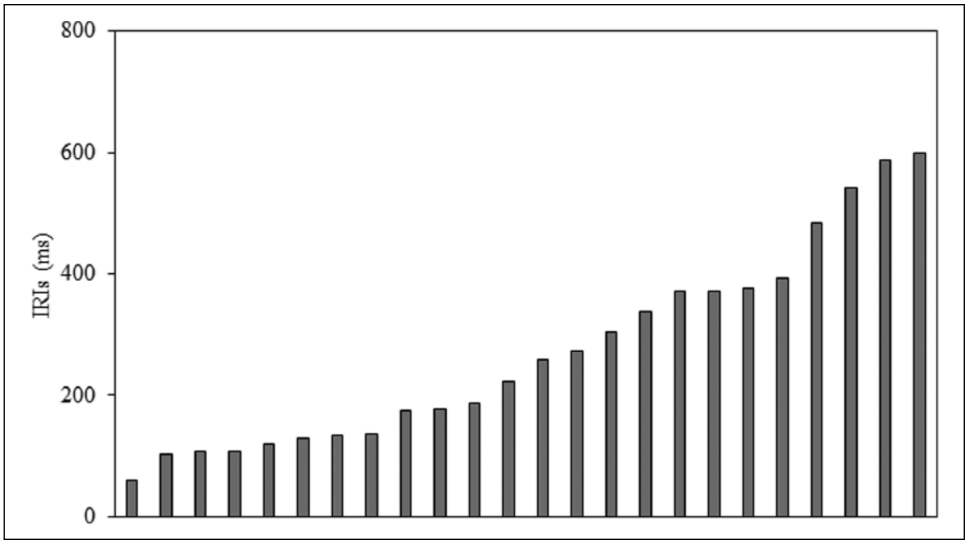

Figure 4. Inter-response interval (IRI) of individual subjects in Experiment 1.

practice phase triads (results were not materially changed when IRIs were computed only over triads 6 through 15, wherein mean RTs did not change significantly). The results are shown in Figure 4, individually ordered by IRI magnitude. In the previous studies, we were able to observe a significant discontinuity in the IRIs at about 300 ms, motivating an approximate grouper vs. non-grouper categorisation as a heuristic to facilitate data analyses as noted earlier: Subjects with mean IRIs of less than 300 ms were classified as *groupers* and subjects with IRIs greater than 300 ms as *non-groupers*. However, as we can see in Figure 4, the data in our current study reflect another pattern that reveals no such discontinuity. Because there is no evidence of a discrete IRI shift, the simple classification of subjects into likely groupers and non-groupers was not motivated. As such, the main analyses below were conducted on all subjects. In the General Discussion, we use an alternative approach to study the subject-level IRI patterns across both studies that compliments the main analysis.

Practice phase dual-retrieval RTs. Figure 5 shows the practice phase means for RT1 and RT2, along with the ES and race predictions across all subjects in each triad. RT1 was above the ES prediction throughout all of the practice triads (Triad 1-15). In contrast to that, RT2 was above the ES prediction during the first triad, fell below the ES prediction on the second and third triad and remained 100 to 150 ms below that prediction throughout the rest of practice (Triad 4-15). Further, t-tests comparing RT2 to the ES prediction separately for each individual triad from Triad 4 on, showed consistently lower RT2s, $ts(23) > -2.989$, $ps < .001$. By the end of the practice phase mean RT2 was at about the mid-point between the ES and race model predictions, and still significantly above the race prediction by the last practice triad, $t(23) = 2.980$, $p < .001$. These results closely match the overall results of the prior experiments on two episodic retrievals.

Linking these findings to those of previous studies strengthens our assumption about the presence of learned retrieval parallelism (Nino & Rickard, 2003; Strobach et al., 2014). However, the present findings go beyond those of previous studies: Whereas the preceding experiments used experimental designs in which subjects had to learn and apply two new and unknown memory retrieval tasks, this experiment used one new and one highly learned memory association which is assumed to help gaining more knowledge about either the unitary- or the multiple memory systems account. Furthermore, since we did not divide the sample into two subgroups (e.g., grouper and non-grouper subjects), we were able to show the phenomenon of learned retrieval parallelism in a complete sample rather than in a specific subgroup in an experiment that used two retrievals from a single cue.

Transfer phase dual-retrieval RTs. Of most interest in the transfer data is performance of subjects on the first dual-retrieval transfer triad (Figure 5). 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; Strobach et al., 2014). There was, however, prior exposure to the dual-retrieval task for other cues during the practice phase, providing an opportunity for subjects to optimise dual-retrieval performance at a global (i.e., non cue-specific) level. As displayed in Figure 5, RT1 was above the ES as well as the race prediction throughout the whole transfer phase. Furthermore, a *t-test* for the first transfer triad with new cues showed that RT2 did not violate the ES lower bound prediction, $t(23) < 1$. In

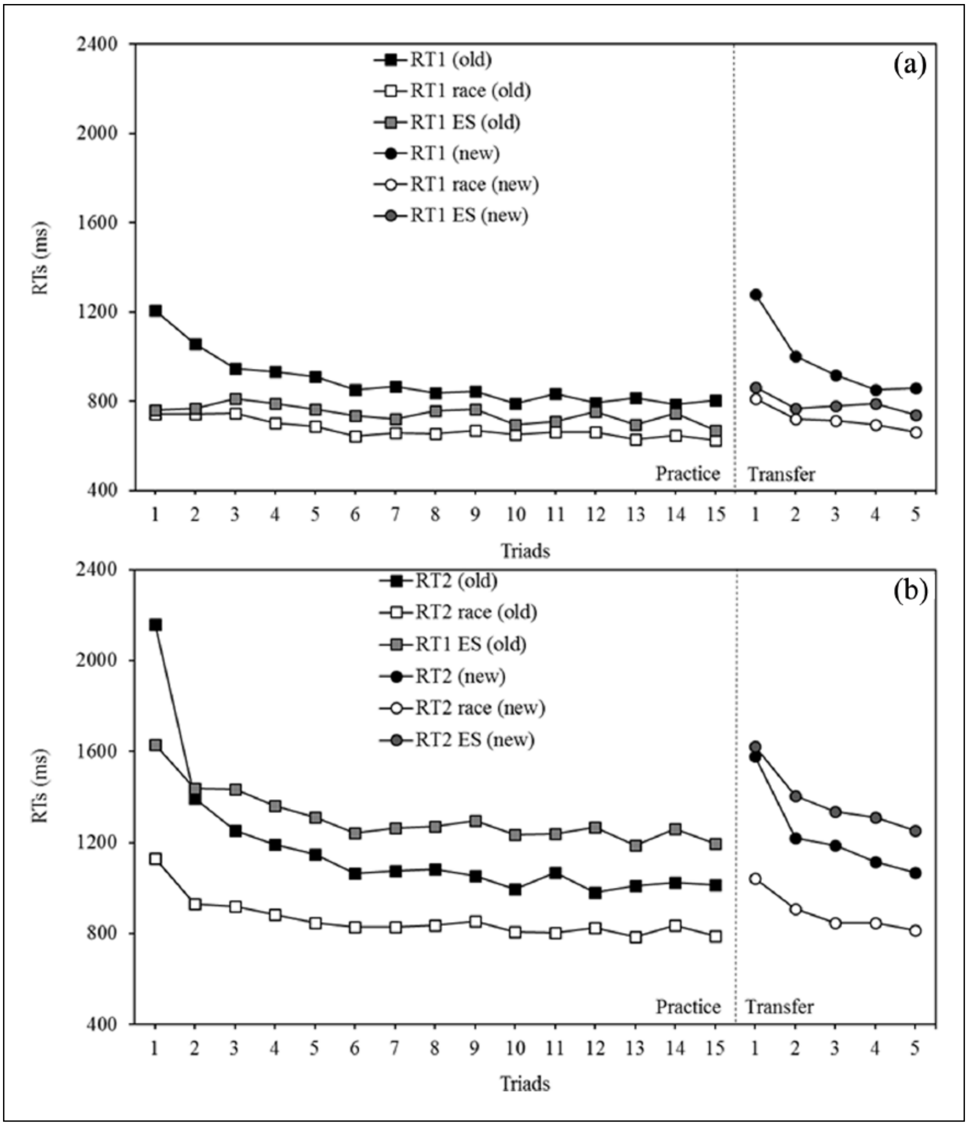

Figure 5. Observed reaction times (i.e., RT1, Panel A and RT2, Panel B) as well the predictions of the efficient sequential (ES) and the race model during the single-dual practice and transfer phases in Experiment 1.

contrast, we found that RT2 fell well below its ES prediction in the final practice triad with old cues (Practice Triad 15), and violated the ES lower bound prediction, $t(23) = -4.267$, $p < .001$. Analyses for RT2 and the race prediction showed that RT2 was above the race prediction by the end of practice, $t(23) = 5.622$, $p < .001$, and remained above this prediction on the first transfer triad, $t(23) = 5.238$, $p < .001$, as well as throughout the whole transfer phase. This pattern was also observed in our prior studies of two episodic retrievals.

In order to support the conclusions above, an analysis of variance (ANOVA) of Triad (limited to Practice Session Triad 15 vs. Transfer Triad 1) and Dataset (RT2 vs. ES vs. race prediction) showed an interaction effect, $F(2, 46) = 68.065$, $p < .001$. The results further exhibited a main effect for Dataset, $F(2, 46) = 57.992$, $p < .001$, along with a significant main effect for Triad, $F(1, 23) = 9.401$, $p < .01$. All of this supports the hypothesis of a shift to sequential retrieval processing for the first transfer triad after practice, and is consistent with the claim that learned parallelism in the practice phase is a cue-level rather than task-level phenomenon.

As an alternative approach to our analyses, we used Bayesian statistics to assess the difference between observed values of RT2 and its associated ES prediction in the first transfer triad. The reason to include Bayesian statistics is based on its strength to create statistical information that is unbiased, has no need for approximation assumptions of the homogeneity of variances and helps to create more informative inferences by reporting combined probabilities of connected parameter values (Kruschke, 2013). Therefore, a Bayesian paired samples t-test was

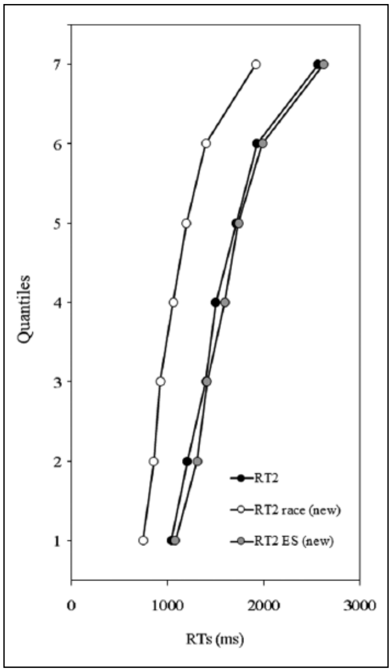

Figure 6. Cumulative distributions of observed reaction times (RT2), as well as the associated race and efficient-sequential (ES) prediction for the first transfer triad in Experiment 1.

conducted in JASP (JASP Version 0.8.0.0; JASP Team, 2017), and we computed its Bayes factor. The Bayes factor (BF_{10}) is the likelihood ratio of the data given the alternative hypothesis against the null hypothesis with larger Bayes factors ($BF_{10} > 1$) being supportive of the alternative hypothesis. For this analysis, we used the default JASP Cauchy prior. The results of our comparison of the first transfer triad's RT2 and its ES prediction indicate moderate evidence for H_0 ($BF_{10} = 0.232$) which strengthens the results of our conclusions based on the null hypothesis test (NHST) analysis which showed that the observed values of RT2 are not statistically different from the ES prediction.

As another approach to more stringently test the compatibly of the RT2 data on the first transfer block with the ES prediction, we calculated the cumulative distribution of RT2 and its associated ES (and race) predictions. Figure 6 shows the plotted results in each of the seven quantiles rank ordered from shortest to longest for each subject. In order to maintain the accuracy and completeness of the rank ordering, we include each response regardless of its accuracy. The cumulative distribution analysis showed that RT2 was numerically almost identical to the ES prediction across all quantiles. Separate t-test performed at each quantile did not approach significance, $ts(6) < 1.31$, $ps > .2$. This cumulative distribution result for RT2 also matches almost exactly that observed for two episodic memory retrievals in Strobach et al. (2014).

Discussion of experiment 1. Overall, the results of this experiment are consistent with the hypothesis that, prior to learned parallelism through dual-retrieval practice, the retrieval stage of processing for an episodic and semantic retrieval are sequential, and that the retrieval bottleneck account and set-cue model are sufficient to explain the data. Combined with our prior results for episodic retrieval, these results further raise the possibility that the bottleneck is a global property of long-term memory retrieval.

Experiment 2

The results of Experiment 1 are consistent with the cue-level chunking account of learned parallelism when we combine an episodic memory-retrieval task (keypress task) with a semantic memory task (the antonym task). By that account, cue-level chunking occurred during dual-retrieval practice but did not exist on the transfer test for cues that had not been previously practised under the dual task. An alternative version of the task-level account, however, still remains plausible: Subjects may have made a strategic shift to parallel retrieval during dual-retrieval practice but then shifted back to sequential retrieval in the transfer phase in the context of new dual-retrieval cues and the absence of old dual-retrieval cues (Strobach et al., 2014). 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 context. We will refer to this account as *context-dependent task-level account.*

In Experiment 2, we therefore 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 from Experiment 1 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 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 the strategy of sequential vs. parallel retrieval should apply to all cues in the transfer phase, 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.

So far, all of the experiments involving the dual-retrieval task (Experiment 1; Nino & Rickard, 2003; Rickard & Pashler, 2005; Strobach et al., 2014) have exclusively been conducted with English word cues and verbal responses (i.e., antonyms, digits). However, language processing and associated semantic processes can

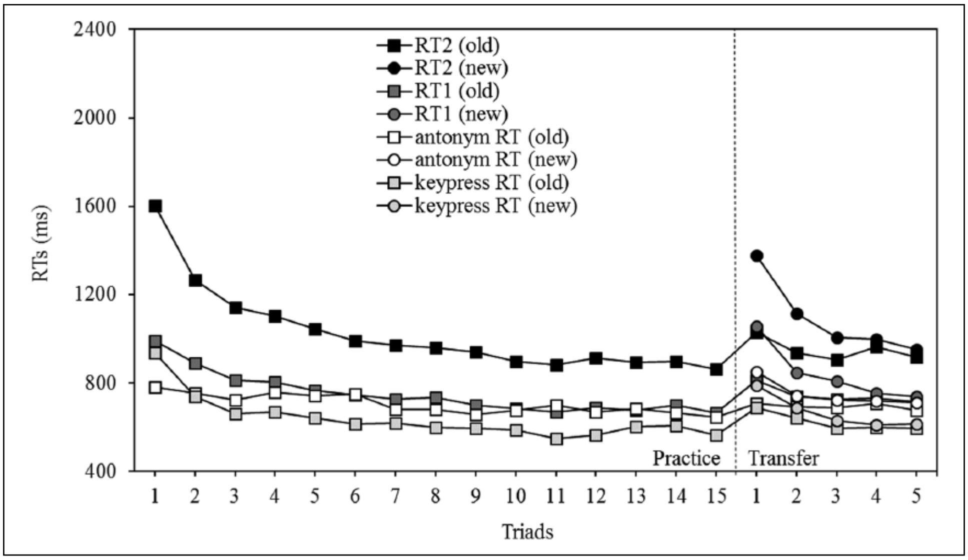

Figure 7. Observed reaction times (RTs) in single-retrieval blocks of the keypress task and the antonym task as well as observed RTs in dual-retrieval blocks (i.e., RT1 and RT2) in the overall dataset during the 15 practice triads and 5 transfer triads in Experiment 2.

differ with regard to associated brain activation, even within the group of Indo-European languages like English and German (Friederici & Chomsky, 2017). This might allow for the possibility that the observed retrieval patterns might not only be cue-specific, but also potentially influenced by language-specific processes. Consequently, since it could provide further insights into previously observed result patterns, we conducted Experiments 2 and 3 using the Germany language with native speaking Germans.

Methods

Subjects. Thirty undergraduate students of the Humboldt University Berlin, Germany, participated. The sample included 19 female and 11 male subjects who had mean age of 24.4 years ($SD = 3.5$). They received course credit or monetary compensation for participation and were naïve of the study's research aim. To avoid an increase in performance motivation, subjects were only able to receive either a predetermined amount of course credit or monetary compensation. This decision was based on research findings which concluded that performance-based compensations are suspected to cause an increase in performance motivation (Bowen & Kensinger, 2017; Brase, 2009). No such significant influence on performance motivation was found for flat-fee monetary compensation and course credit (Brase, 2009). All subjects were German native speakers, right-handed and had normal or corrected-to-normal vision.

Apparatus, cues, design, and procedure. All of the elements were identical to Experiment 1 with the following exceptions. The list of German cues, the corresponding keypress responses in the keypress task, and the German antonym word responses in the antonym task are shown in Table 1. In contrast to the single–dual transfer phase in Experiment 1, each block included the presentation of all 14 cues (i.e., *old cues* practised during the single–dual practice phase and *new cues* that were introduced before this phase but were not practised here) and thus 14 trials (Figure 2).

Results and discussion

Accuracy results. In all analyses, we excluded trials with RTs below 200 ms (0.4% of all single- and dual-retrieval trials). In the single–dual practice phase (i.e., from Practice Triad 1 to Practice Triad 15), error rates decreased from 5.5% to 1.2% for the keypress single-retrieval trials, and from 2.4% to 0.5% for the antonym single-retrieval trials. For the keypress task in the dual-retrieval blocks of this phase, error rates decreased from 11.9% to 1.2%, and for the antonym task from 4.5% to 0.3%. Dual-retrieval error rates during the single-dual practice phase decreased from 7.2% to 0.8% for the first completed response (keypress or antonym), and from 9.3% to 0.2% for the second response.

RT results. RTs averaged over all subjects for correctly performed single-retrieval (i.e., keypress task, antonym task) and for dual-retrieval trials (RT1, RT2) are shown in Figure 7. RTs decreased steadily over the course of single-dual practice (Practice Triad 1-15), but increased markedly on the transfer test, particularly for dual-retrieval trials, on new cues and Transfer Triad 1.

IRI analysis. Following analyses in Nino and Rickard (2003) as well as Strobach et al. (2014), we computed individual

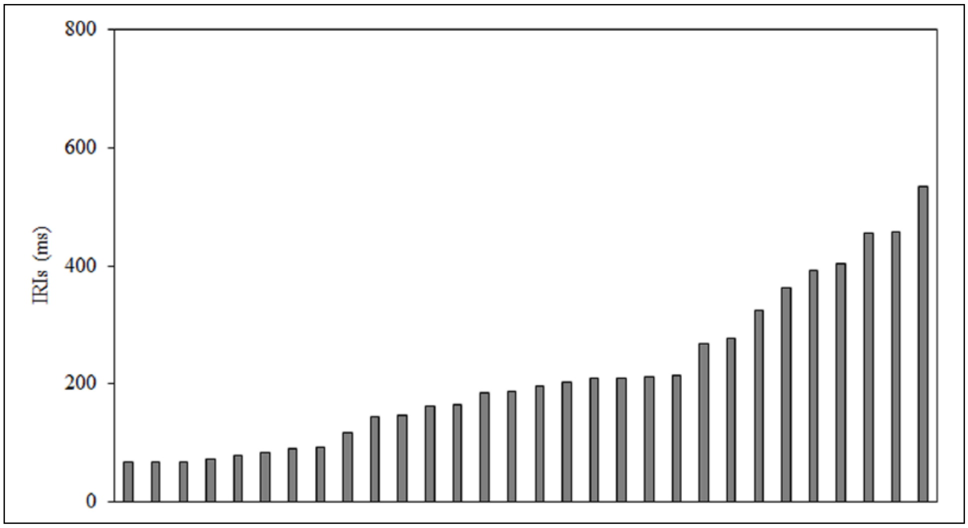

Figure 8. Inter-response interval (IRI) of individual subjects in Experiment 2.

mean IRIs on dual-retrieval trials for each subject, averaging over all practice phase triads. The results are shown in Figure 8, individually ordered by IRI magnitude. Similarly, to Experiment 1, the data of Experiment 2 reveals a rather continuous pattern of the IRI distribution. Thus, we again did not divide the subjects into different categories (e.g., grouper and non-grouper subjects), but rather performed overall analyses.

Practice phase dual-retrieval RTs. Figure 9 displays the practice phase means for RT1 and RT2, along with the ES and race predictions across all subjects in each triad. What we can observe in Panel A of this figure is that RT1 was constantly above the ES as well as above the race prediction throughout the practice phase. In contrast to that, RT2 (Panel B) was above the ES prediction during the first triad, aligned the ES prediction on the second triad, fell slightly below on the third triad and remained 100-150 ms below that prediction throughout the rest of practice. Looking at *t*-tests comparing RT2 to the ES prediction after each individual triad from Triad 3 on, showed lower RT2s for each triad, $ts(29) > 2.180$, $ps < .05$, except for Triad 4, $t(29) < 1$, and Triad 9, $t(29) < 1$; with a strong violation of the ES prediction in the last practice triad, $t(29) = 3.688$, $p < .001$. Based on findings from Experiment 1 as well as from diverse prior findings, the ES model provides an empirically validated reference prediction for the evaluation of learned retrieval parallelism among subjects. These results clearly point to some form of learned retrieval parallelism, extending findings of previous studies (Experiment 1; Nino & Rickard, 2003; Strobach et al., 2014) to German native speakers.

Transfer phase dual-retrieval RTs. Similar to the previous experiment, the performance for subjects on the first dual-retrieval transfer triad is of most interest for the questions under investigation. As shown in Panel A of Figure 9, RT1 for old cues was above the ES and the race prediction. The same account holds for the RT1 values for new cues. While RT2 for old cues was, as expected, below its ES prediction during this first transfer triad, $t(29) = -3.561$, $p < .001$, there was no evidence for a statistical difference between RT2 and the ES prediction on new cues, $t(29) < 1$ (Figure 9b). In addition, RT2 remained above the race prediction for old cues, $t(29) = 6.812$, $p < .001$, as well as new cues, $t(29) = 5.689$, $p < .001$, on the first transfer triad. This result pattern was further qualified by a repeated measures ANOVA of Cue type (old cues vs. new cues) and Dataset (RT2 vs. ES vs. race prediction). That analysis showed a significant interaction, $F(1, 58) = 5.081$, $p < .01$, as well as a main effect for Dataset, $F(2, 58) = 56.460$, $p < .001$, and for Cue type, $F(1, 29) = 42.207$, $p < .001$. After the first transfer triad, RT2 for new cues fell below the ES prediction for the rest of the transfer phase (Transfer Triad 2-5), $ts(29) > -2.362$, $ps(29) < .001$. In addition, RT2 for new cues remained above the race prediction throughout the transfer phase, $ts(29) > 5.402$, $ps < .001$ (Transfer Triad 2-5). We can conclude from these results (mainly from the first transfer triad) that they are in line with the hypothesis of learned retrieval parallelism at the cue level and not with the context-dependent task level. These training and transfer results also closely parallel results of Experiment 2 of Strobach et al. (2014) for two episodic retrievals.

Likewise to Experiment 1, we additionally computed the Bayes factor for the difference of RT2 and the ES prediction in the first transfer triad. This analysis was conducted separately for old and new cues. The Bayesian paired samples *t*-test for old cues indicated a strong evidence for the alternative (H_A) hypothesis that RT2 and the ES prediction differ from each other ($BF_{10} = 26.355$; Cauchy prior = .707). In contrast, the results for new cues indicate moderate evidence for H_0, meaning that RT2 and

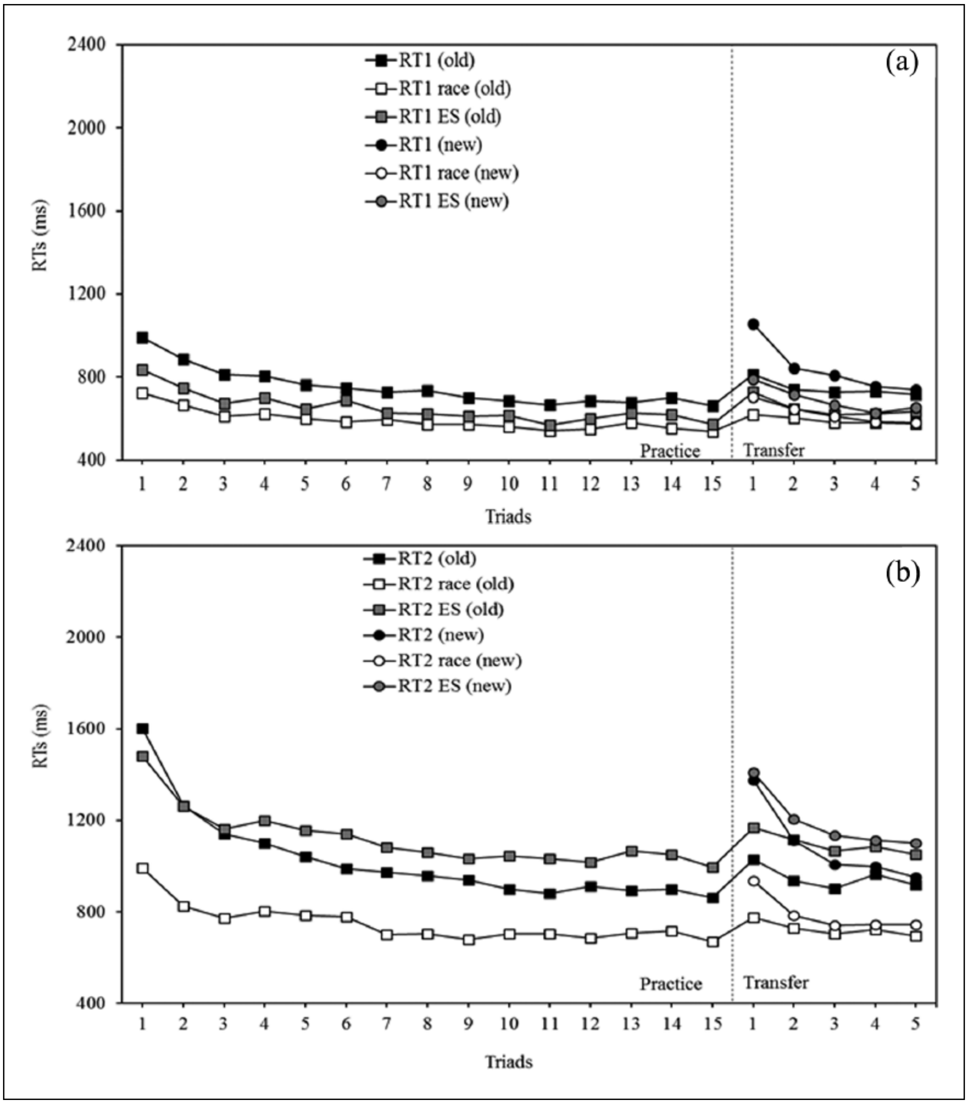

Figure 9. Observed reaction times (i.e., RT1, Panel A and RT2, Panel B) as well the predictions of the efficient sequential (ES) and the race model during the single-dual practice and transfer phases in Experiment 2.

the ES prediction are moderately likely not to differ from each other (BF_{10} = .21). These results are in line with the results of our NHST analyses stated above and further support the hypothesis of cue-level learned retrieval parallelism.

As in the first experiment, we computed and plotted the cumulative distributions for all valid responses for RT2, and its associated predictions, in the first transfer triad (Figure 10). These analyses were conducted separately for old (Panel A) and new cues (Panel B) across all subjects. For the old cues, we observed a significant difference between RT2 and its ES prediction on all quantiles, $ts(29) > -3.865$, $ps < .02$, except for Quantile 7, $t(29) = -1.733$, $p > .5$. For the new cues, RT2 was not statistically different from the ES prediction across quantiles 2 to 7, $ts(29) < -1.690$, $ps > .1$. The only exception was formed by Quantile 1 which was significantly different from the ES prediction, $t(29) = -2.315$, $p < .05$. These results are, again, largely consistent with the cue-specific account of learned parallelism.

Experiment 3

In Experiments 1 and 2, we provided evidence for the cue-specific account of learned parallelism in situations with a novel (episodic) keypress task and the automatised (semantic) antonym task. So far, there has been no evidence for learned parallelism at the task level. However, that result may stem from an insufficient amount of dual-retrieval practice. With an increase in the practice dosage, subjects may acquire skills that generalise from old cues to new cues and thus, lead to a shift from cue-specific learned retrieval parallelism to task-general learned parallelism, such as indicated by studies on dual-choice RT practice

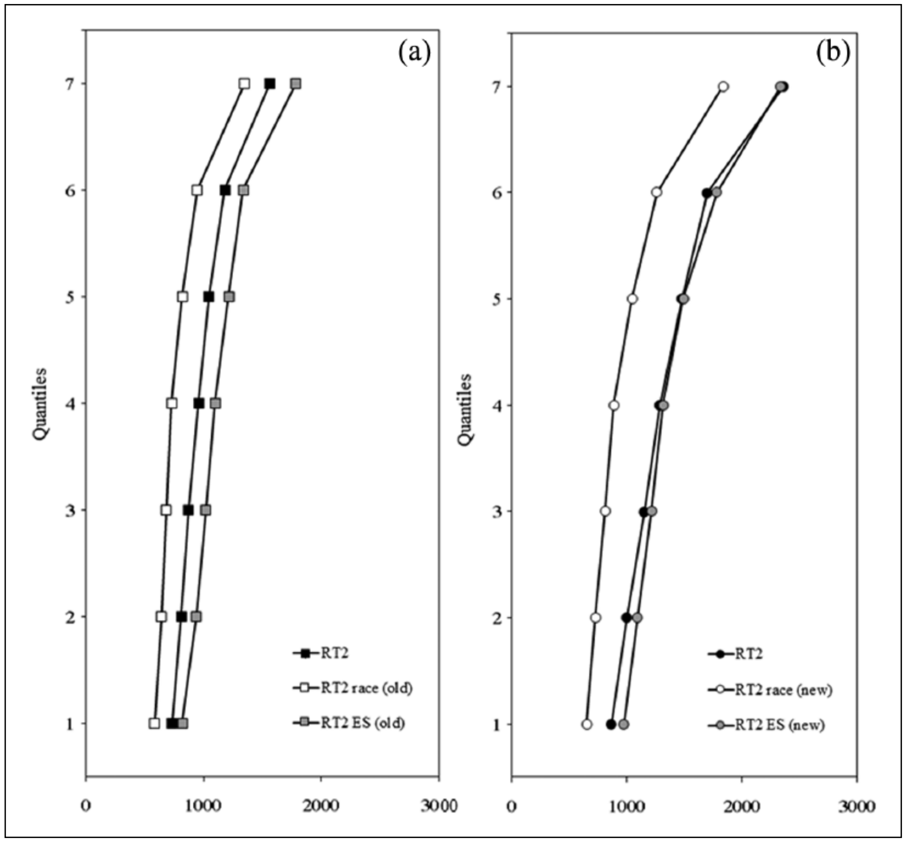

Figure 10. Cumulative distributions for old (Panel A) and new (Panel B) cues of observed reaction times (RT2), as well as the associated race and efficient-sequential (ES) prediction for the first transfer triad in Experiment 2.

(Maquestiaux et al., 2008; Ruthruff et al., 2006). Such a dose-dependence was also demonstrated in a different setting: the acquisition of general skills that transfer to unpractised tasks on reasoning abilities after working-memory training is dose-dependent. For example, there is no transfer after 8 and 12 working-memory practice sessions, but after 17 and 19 sessions (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; but see Redick et al., 2013). As a consequence of the possible impact of a dose-dependence on transfer effects, we aimed to test the generality of learned retrieval parallelism and the set-cue bottleneck account after an increased amount of dual-retrieval practice with German antonyms. Thus, after we provided 15 practice triads in Experiments 1 and 2, we increased this number to 40 triads for German native speakers in Experiment 3.

Methods

Subjects. Thirty-six undergraduate students of the Humboldt University Berlin, Germany, participated. The sample included a total of 14 male and 22 female subjects who had a mean age of 23.7 years (SD = 3.9). They received course credit or monetary compensation for participation and were naïve of the study's research aim. All of the subjects were German native speakers, had normal to corrected-to-normal vision and were right-handed.

Apparatus, cues, design, and procedure. These elements were identical to Experiment 2 with the following exceptions. Session 2 ended with 20 triads of the single–dual practice phase. Another 20 triads of this phase (resulting in 40 triads in the single-dual practice phase) preceded the single–dual transfer phase (5 triads) in Session 3 (Figure 2).

Results and discussion

Accuracy results. In all analyses, we excluded trials with RTs below 200 ms (0.5% of all single- and dual-retrieval trials). In the single–dual practice phase (i.e., from Practice Triad 1 to Practice Triad 40), error rates decreased from 6.8% to 0.7% for the keypress single-retrieval trials, and from 5.3% to 1.4% for the antonym single-retrieval trials. For the keypress task in the dual-retrieval blocks of this phase, error rates decreased from 9.9% to 0.9%, and for the antonym task from 6.2% to 0.6%. Dual-retrieval error rates over the course of the single–dual practice phase decreased from 7.8% to 0.8% for the first completed response (keypress or antonym), and from 10.2% to 0.8% for the second response.

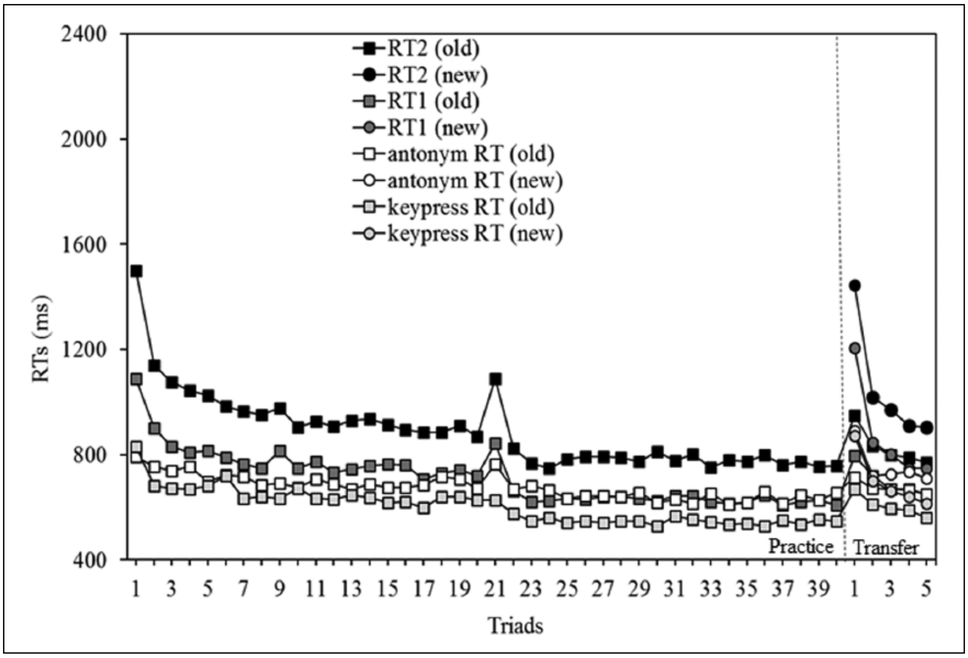

Figure 11. Observed reaction times (RTs) in single-retrieval blocks of the keypress task and the antonym task as well as observed RTs in dual-retrieval blocks (i.e., RT1 and RT2) in the overall dataset during the 40 practice triads and 5 transfer triads in Experiment 3. Increased RTs in Triad 11 reflect the start of Session 3 after the end of Session 2 (Triad 21).

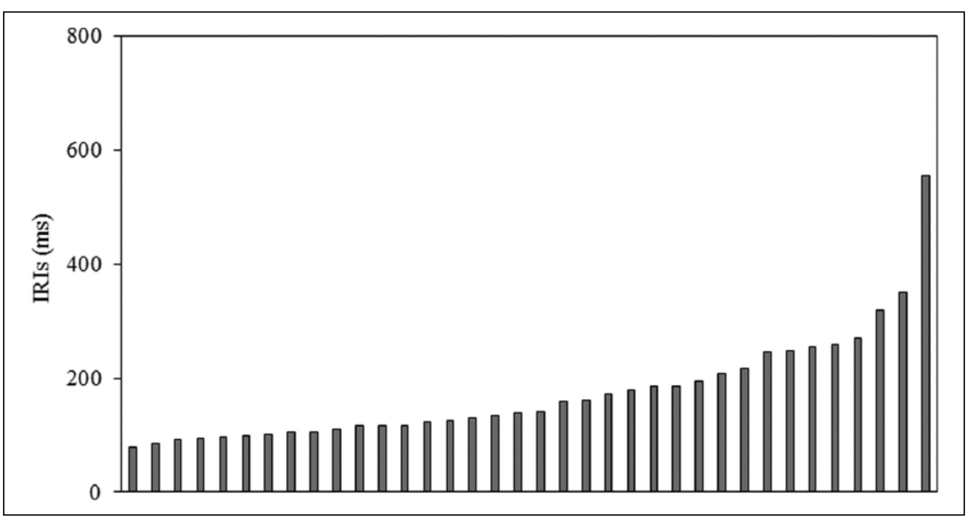

Figure 12. Inter-response interval (IRI) of individual subjects in Experiment 3.

RT results. RTs averaged over all subjects for correctly performed single-retrieval (i.e., keypress task, antonym task) and for dual-retrieval trials (RT1, RT2) are shown in Figure 11. RTs decreased steadily over the course of single-dual practice (Practice Triad 1-40), but increased markedly on the transfer test, particularly for dual-retrieval trials, new cues, and Transfer Triad 1. Increased RTs in Practice Triad 21 mark the restart of the single–dual practice phase at the beginning of Session 3, and presumably reflect forgetting effects over the between-session delay.

IRI analysis. Like in the previous experiments, we computed individual mean IRIs on dual-retrieval trials for each subject, averaging over all practice phase triads (Figure 12). The data reflect the same continuously ascending pattern that we observed in Experiments 1 and 2 as well which accounts for our repeated decision not to divide the subjects into sub-groups or categories (e.g., grouper and non-grouper subjects). Therefore, we performed analyses across the whole sample like we did in the two prior experiments in this paper.

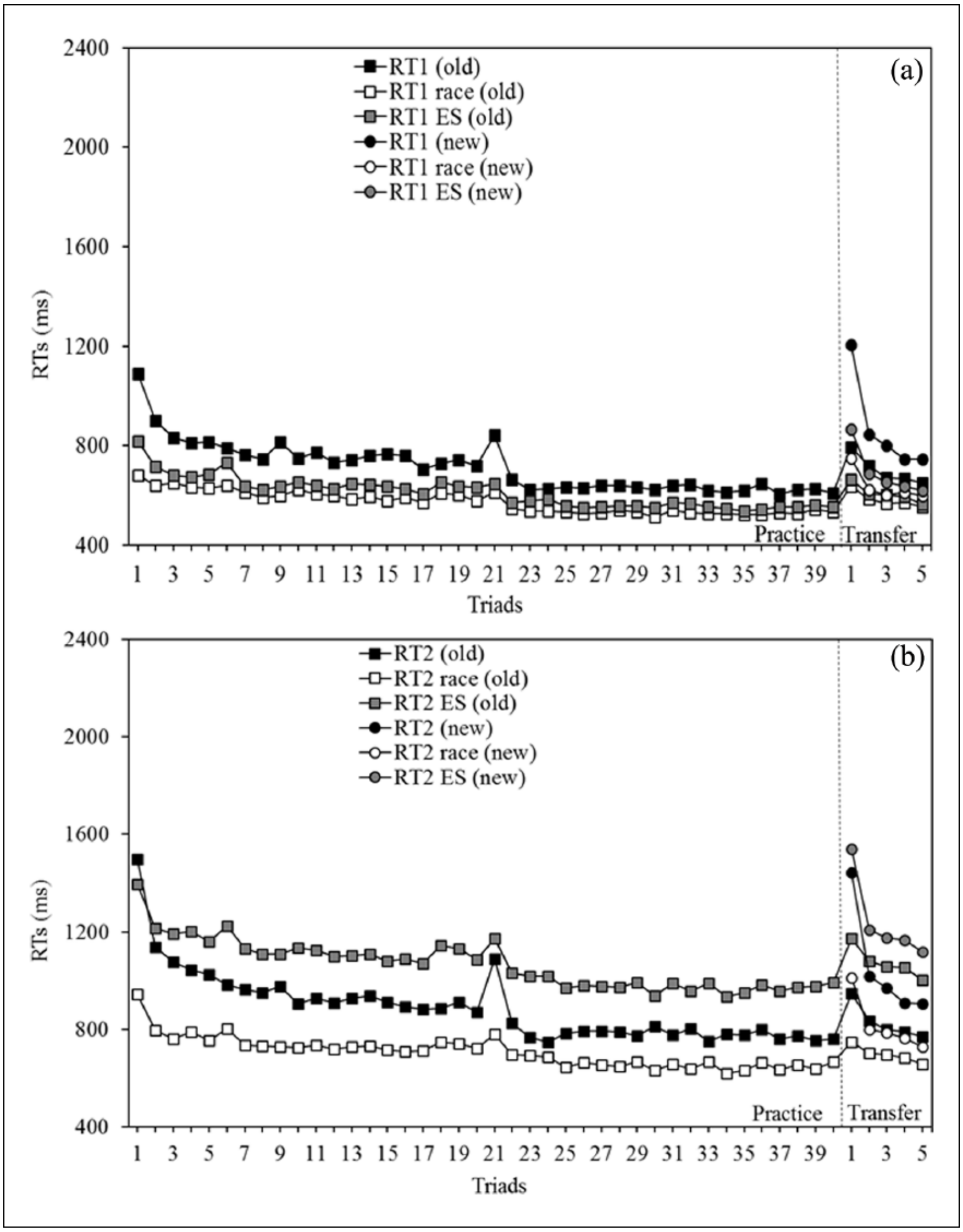

Figure 13. Observed reaction times (i.e., RT1, Panel A and RT2, Panel B) as well the predictions of the efficient sequential (ES) and the race model during the single-dual practice and transfer phases in Experiment 3. Increased RTs in Triad 21 reflect the start of Session 3 after the end of Session 2 (Triad 21).

Practice phase dual-retrieval RTs. The practice phase means for RT1 and RT2 across all subjects in each triad are displayed in Figure 13. Results for the ES as well as the race predictions are presented as well. As in the previous experiments, RT1 is consistently located above the ES and the race prediction throughout the whole practice phase. The first triad of the practice phase shows that RT2 was located approximately 100 ms above the ES prediction, but it fell below this prediction from the second triad on. t-tests comparing RT2 to the ES prediction for each individual triad after Triad 1, showed lower RT2s, $ts(35) > 2.113$, $ps < .05$, except for Triad 21, $t(35) < 1$. Encompassing the previous findings in this experiment as well as others (Nino & Rickard, 2003; Strobach et al., 2014), we can conclude that these results indicate a form of learned retrieval parallelism in an experiment with extended practice.

Transfer phase dual-retrieval RTs. As we mentioned in the previous experiments, of most interest in the transfer data is the performance of subjects on the first dual-retrieval transfer triad. As we can observe in Panel A of Figure 13, RT1 was located above the ES and the race prediction for old and for new cues for the first as well as throughout all transfer triads. Whereas RTs violated the ES prediction for

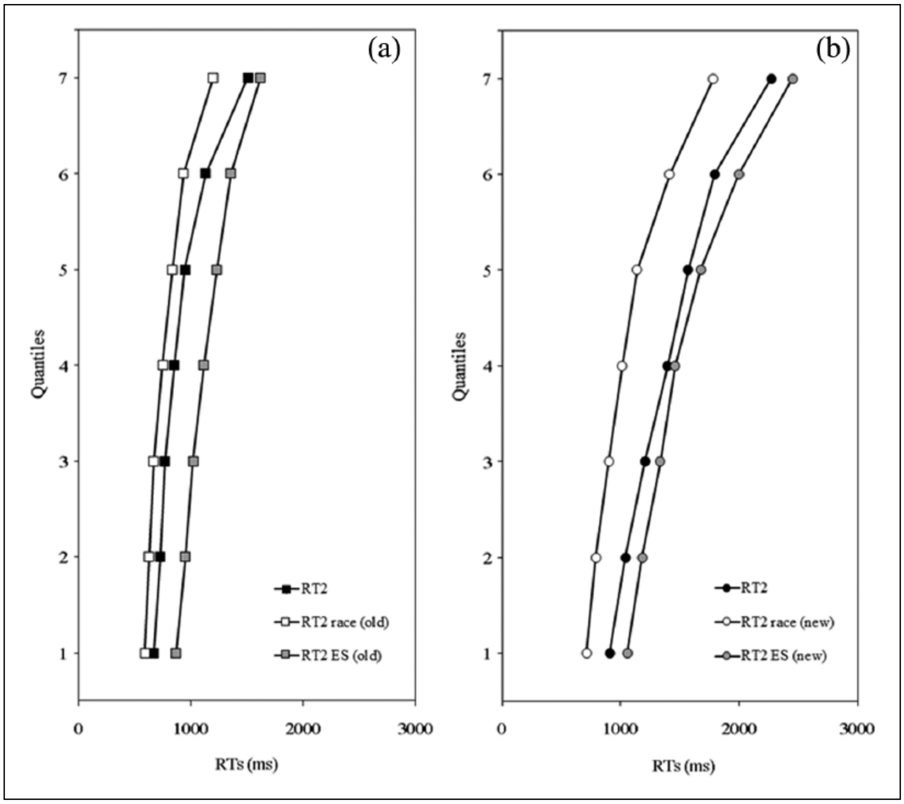

Figure 14. Cumulative distributions for old (Panel A) and new (Panel B) cues of observed reaction times (RT2), as well as the associated race and efficient-sequential (ES) prediction for the first transfer triad in Experiment 3.

old cues in the first transfer triad, $t(35) = -8.535$ $p < .001$, there was no evidence for that statistically for new cues, $t(35) < 1$. In contrast, RT2 was above the race prediction for old, $t(35) = 6.886$, $p < .001$, as well as for new cues, $t(35) = 4.028$, $p < .001$. These effects were displayed by a significant interaction effect in a repeated measures ANOVA on the factors Cue type (old cues vs. new cues) and Dataset (RT2 vs. ES vs. race prediction), $F(2, 70) = 3.610$, $p < .05$. As in Experiment 2, RT2 for new cues fell significantly below the ES prediction for the rest of the transfer phase (Transfer Triad 2-5), $ts(35) > -6.185$, $ps < .001$, but remained above the race prediction, $ts(35) > 4.003$, $ps < .001$. Again, we can conclude that these findings are consistent with assumptions of learned retrieval parallelism at the cue level account.

As in the previous experiments, we computed the Bayes factor for the difference of RT2 and the ES prediction in the first transfer triad, separately for old and new cues. Extending the results from Experiment 2, the Bayesian paired samples *t*-test for old cues indicated decisive evidence for the alternative (H_A) hypothesis that RT2 and the ES prediction are statistically different from each other ($BF_{10} = 2.430e + 7$). In addition, the results for new cues indicate moderate evidence for H_0, which showed that RT2 and the ES prediction are moderately likely not to differ ($BF_{10} = .256$). As concluded before, these results are in line with the NHST analyses and support our hypothesis of a cue level account in learned retrieval parallelism.

In line with the previous two experiments, we plotted the cumulative distributions of RT2 and both predictions for the first transfer triad across all subjects (Figure 14). As in Experiment 2, these analyses were completed separately for old (Panel A) and new cues (Panel B). For the old cues, the observed values for RT2 and its associated ES prediction significantly differed across quantile 1 to 6, $ts(36) > -10.807$, $ps < .001$. In contrast, RT2 for the new cues did not differ significantly from the ES prediction across quantiles 3 to 7, $ts(36) < -.535$, $ps > .126$, though it did differ significantly for the first two quantiles. Overall, these results largely support the cue-specific account of learned parallelism, with the possible caveat of low order distribution quantiles after extensive practice in Experiment 3.

General discussion

The central aim of this study was to determine whether the prior conclusions about dual memory retrieval of two novel associations from a single cue (Strobach et al., 2014) extend to the case of one novel and one automatised memory retrieval. For Experiments 1 and 2, the major empirical results were highly similar to those of two closely related prior experiments that explored two novel, and episodic, retrievals (Nino & Rickard, 2003; Strobach et al., 2014). In both paradigms, the mean values of RT2 were longer than predicted by the ES model on the first practice trial. During the course of practice, dual-retrieval yielded mean RT2 values that were substantially below the ES prediction. Finally, on the first block of the transfer phase in both paradigms, RT2 for new cues (i.e., cues for which only single retrievals had been practised) was again consistent with the ES model prediction, suggesting a reversion back to sequential retrieval stage processing. Those results occurred when only new cues were included in the transfer phase (Experiment 1) and when both old and new cues were randomly intermixed in the transfer phase (Experiment 2). Thus, in nearly all major respects (with exceptions discussed below), the current results were highly analogous to the results for two novel and episodic retrievals (e.g., Strobach et al., 2014). Furthermore, the current results generalise our findings to the German language.

We interpret those results in terms of the set-cue bottleneck model, which assumes a retrieval stage processing bottleneck plus a cue-level response chunking process that occurs with dual-retrieval practice for the cases of both two novel episodic retrievals and for the combination of one novel and one automatised retrieval. We further infer from these results that the retrieval stage processing bottleneck might be a global memory mechanism since we were able to observe it in a new retrieval context of one episodic and one semantic retrieval.

The cue-level chunking process appears to depend on the synchronised presence of both responses in working memory, which we hypothesise, allows the chunking mechanism to operate and learned retrieval parallelism to occur. In addition, the assumption that cue-level chunking can eliminate a working memory bottleneck is further supported by the results of this study, since the potential bottleneck-like focus of attention is not only eliminated for two episodic memory retrievals, but also for a combination of episodic and semantic retrievals. Moreover, the cue-level chunking account also predicts the current finding of minimal evidence for parallel retrieval for new dual-retrieval cues on the transfer test (with the partial exception of Experiment 3 as discussed below). Importantly, learned retrieval parallelism in the chunking account does not require abandonment of the retrieval stage bottleneck hypothesis. Rather, we assume that the chunking process allows both responses to be retrieved in a single pass through the bottleneck, and depicted for the set-cue model in Figure 1.

Implications for the generality of the proposed dual-retrieval bottleneck model

Our results add to a growing body of evidence that, contrary to some interpretations of the *multiple memory systems* account, the retrieval of semantic and episodic information may depend on similar processes. The most natural prediction within the *multiple memory systems* framework would appear to be that parallel retrieval through episodic and semantic memory should involve no interference, and no processing bottleneck, within each respective memory system (García-Lázaro et al., 2012; Rajah & McIntosh, 2005). Because our task also involves zero or very low interference in the perceptual and motor processing stages, we would expect by that account that parallelism would be observed prior to dual task learning (i.e., on the first dual-retrieval training block). Our results instead seem more compatible with the *unitary memory system* account, although it should be acknowledged that the *unitary memory system* account was not developed to specifically address the issue of parallel vs. sequential retrieval processes (Rajah & McIntosh, 2005).

The results of Experiment 3, which involved substantially greater single–dual retrieval practice, were also similar to those of prior experiments (Nino & Rickard, 2003; Strobach et al., 2014), with the exception being that, uniquely in that case, there was some evidence of task level parallelism. That pattern was apparent by all measures for new cues on the transfer test, but most notably for the lower cumulative distribution quantiles in Figure 13b. We tentatively infer that if the same extensive practice were given for the case of two episodic retrievals, a similar degree of task-level parallelism would be observed. That possibility could be pursued in further investigations, since different task-sets that involve altered retrieval processes could account for the facilitation of task level parallelism instead of cue-specificity. Moreover, it could be speculated that the partial evidence for task level parallelism in Experiment 3 could be due to inaccurate lower-bound rates predicted by the ES model. Therefore, even though previous examinations of the ES prediction showed that it indeed rather accurately predicts the observed RT2 values for strict sequential processing (Nino & Rickard, 2003; Strobach et al., 2014), there could be a chance for improvements to better reflect the lower bound rate for sequential processing.

Investigation of differences in IRI patterns in the current vs. prior experiments

In both Nino and Rickard (2003) and Strobach et al. (2014), subjects could be reasonably classified as either synchronising responses on most or all trials (yielding short mean IRIs; response groupers) or as executing each response as soon as it was retrieved (yielding long mean IRIs; non-groupers). That method of estimating grouper vs. non-grouper classification was not viable in the current experiments, however, as there were no discrete gaps in the IRI plots through which a discrete classification could be approximated. As such, our primary analyses above were conducted on the full set of subjects only, and the more fine grained comparison of individual differences in strategy selection of prior studies was not performed. In this section, we address that limitation by adopting an alternative approach to exploring individual differences in response strategy and learned parallelism that can be applied consistently to both the current data and the Strobach et al. data.

The first step in this alternative approach is to identify subjects for whom mean RT2 towards the end of practice was consistent with the possibility of sequential retrieval, versus subjects for which RT2 towards the end of practice was most consistent with parallel retrieval. Specifically, if a subject's mean RT2 was greater than or non-significantly different from the ES model's RT2 prediction, then that subject was classified as a candidate sequential retriever. Alternatively, if a subject's mean RT2 was significantly below the ES prediction (at alpha = .05), then the subject was classified as a candidate parallel retriever. Although this classification is unlikely to be exact, it serves the purpose of providing a matched and more fined grained comparison to the prior experiments involving dual episodic retrieval. In each experiment from both the current and prior study, those calculations were limited to the last nine single–dual practice triads over which grand mean RT2 did not change substantially, indicating that subjects were approaching at least a within-session performance asymptote. The second step was to indicate, on the subject-level IRI plots for each experiment, which subjects were classified as sequential retrievers and which as parallel retrievers by the criterion described above.

Results over all experiments in both the current study and in Strobach et al. are shown in Figure 15. Panels A, D, G, J, and M show the classification into sequential and parallel retrievers (for each study and experiment) in relation to subject-level mean IRIs. Other than the lack of discrete IRI shifts in the current study, the patterns for the two studies are quite similar: subjects with relatively long IRIs tended to have RT2 values that are consistent with sequential retrieval, whereas subjects with short IRIs tended to have RT2 values that are consistent with parallel retrieval. Furthermore, Panels B, C, E, F, H, I, K, L, N, and O show observed RT2 and the ES prediction of sequential retrievers as well as parallel retrievers in the last nine practice triads and the first transfer triad. Similar across all five experiments, RT2 is similar to its ES prediction for subjects classified as sequential retrievers and below this prediction for subjects classified as parallel retrievers.

Regarding the unique lack of a discrete IRI shift in the current experiments, one possible account is that dual-retrieval from episodic and semantic memory involves fundamental different properties than do two episodic retrievals, yielding no discrete shift. We cannot advance a more specific process hypothesis along those lines, however. An alternative that seems more likely in our view is that for at least some subjects, the response strategy in the current experiments was more mixed over cues than is the case for two episodic retrievals, eliminating the discrete IRI shift over subjects in the averaged IRI data. That speculation is based primarily on the fact that antonym retrieval in the current experiments is highly overlearned, and thus subjects are likely to have high confidence in correct retrieval, regardless of their response strategy. In contrast, high confidence is unlikely for newly learned episodic memory. Hence, a strict, consciously maintained strategy of always grouping or non-grouping may be less common in the current experiments. It appears to also be the case in Figure 15 that more subjects achieved parallel retrieval in the current experiments, particularly in Experiments 1 and 3. Again, that result may be related to higher retrieval confidence in the antonym task, whereas sequential retrieval is arguably the safer and less risky option for retrieving correct responses (e.g., Fischer & Plessow, 2015). If one task is considered highly likely to yield a correct response, it might decrease the perceived risk of mistakes, promoting response grouping on some trials for subjects that otherwise would not group responses.

Figure 15 also reveals a pattern that, although evident in both studies, was unexpected based on the prior approach to IRI analyses. Namely, across most experiments, several subjects who exhibited short IRIs had RT2 results that were consistent with sequential retrieval. That finding suggests that response grouping, which should yield short IRIs, either is not a sufficient condition to yield parallel retrieval, or that it *is* a sufficient condition, but for some subjects there was not enough dual-retrieval practice for learned parallelism to occur. We direct the reader to Panel M of Figure 15 (data from Experiment 3 of the current study) for evidence that is consistent with the later conclusion. In that experiment—which involved more extensive dual-retrieval practice than any of the other experiments, all but one subject with short IRIs exhibited evidence of parallel retrieval, and all subjects with very short IRIs (less than about 150 ms) exhibited evidence of parallel retrieval. Hence, we tentatively conclude that grouped responding is, given enough dual-retrieval practice, likely to be sufficient to produce parallel retrieval. We further speculate that grouped responding may be necessary, or at least highly conducive, to achieving parallel retrieval; in only one case did RT2 fall below the ES prediction during the practice phase when the mean IRI was relatively long and potentially consistent to the non-grouped responding (see Figure 15a, third bar from the right).

In sum, while recognising that these post hoc analyses cannot support strong causal inference, they do suggest two conclusions that are important for this line of work: (1) there is little evidence that the dynamics of dual-retrieval and learned parallelism differ substantially for the cases of two novel (i.e., episodic) retrievals vs. one novel (i.e., episodic) retrieval and one automatised (i.e., semantic memory) retrieval, and (2) the apparent strategy choice

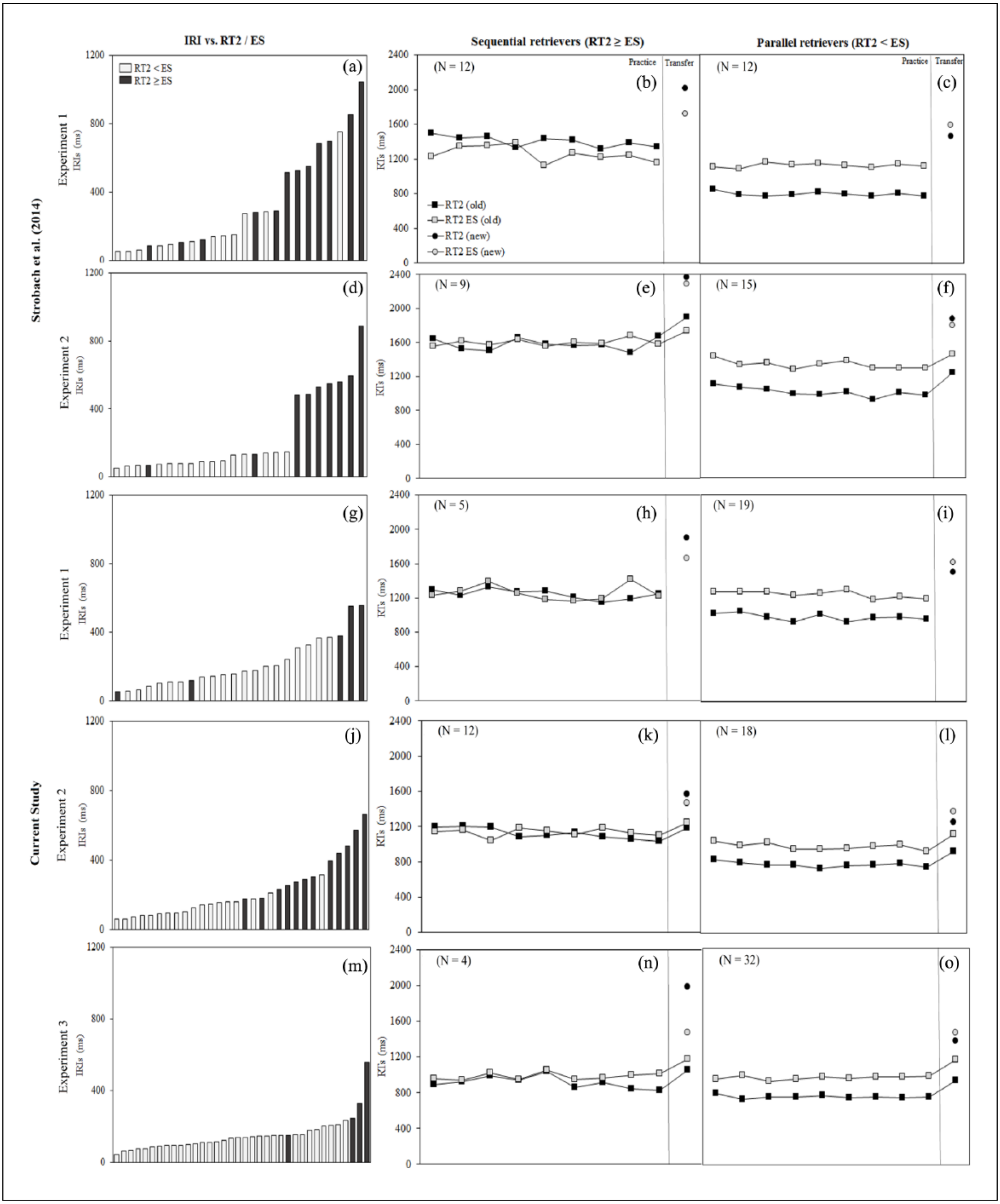

Figure 15. Panels A, D, G, J, and M display the inter-response interval (IRI) of individual subjects in Experiment 1 and 2 by Strobach et al. (2014) (Panel A and D) as well as in Experiment 1 to 3 of the antonym data (Panels G, J, M). Black bars indicate subjects that exhibited sequential retrieval, whereas white bars indicate parallel retrieval. Panels B, E (Strobach et al., 2014, Experiment 1 and 2, respectively), H, K, and N (present Experiment 1, 2, and 3, respectively) exhibit observed reaction times for the second responses (RT2) as well as the predictions of the efficient sequential (ES) model across the last nine practice triads and the first transfer triad for each experiment in the dual-retrieval task for sequential retrievers. Accordingly, Panels C, F (Strobach et al., 2014, Experiment 1 and 2, respectively), I, L, and O (present Experiment 1, 2, and 3, respectively) display the observed RT2 values as well as the ES predictions for the parallel retriever subgroup.

to either group both responses in close proximity, or not, is a critical driver of learned parallelism in two retrievals from a single cue, across both memory systems and both languages studied to date.

Comparison with previous studies on dual-task practice

There exist findings from practice in other dual-retrieval situations that might be consistent with chunking at the task level, rather than the cue level as appears to be the case in the current tasks (with the possible exception of Experiment 3). In situations involving two choice RT tasks (Schumacher et al., 2001; Strobach, Frensch, Müller, & Schubert, 2012a, 2012b, 2015) for example, exclusive single-task practice on some cues results in final dual-retrieval performance that is as good as that achieved through mixed dual- and single-task practice on other cues (Hazeltine, Teague, & Ivry, 2002). The authors interpreted their results to reflect parallel and automatised central response-selection stage processing for the two tasks without the assumption of a bottleneck process at the end of practice (Schumacher et al., 2001; Strobach et al., 2012a, 2012b, 2015). 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, Taatgen, & Byrne, 2005; Ruthruff, Johnston, Van Selst, Whitsell, & Remington, 2003; Schubert, 2008; Strobach & Schubert, 2017). In that account, the two choice RT tasks are processed independently and unchunked, but the structural bottleneck characteristic remains (cf. Nino & Rickard, 2003) making this bottleneck assumption a very general phenomenon when two tasks of the same type are combined (i.e., two memory retrieval tasks or two choice RT tasks). This latter bottleneck assumption is also consistent with the conjoint suggestions of a response selection bottleneck at low practice levels (e.g., Pashler, 1994, 1998) that do not change qualitatively with practice (Nino & Rickard, 2003).

However, this bottleneck assumption at low practice levels might not account for combinations of different task types. That is, when memory retrieval and choice RT tasks are combined in PRP dual-task situations (e.g., Pashler, 1994; Schubert, 1999), there is evidence for parallel processing (Fischer, Miller, & Schubert, 2007; Thomson, Watter, & Finkelshtein, 2010). In particular, retrieval of semantic information of a second component task in PRP situations influences the effectiveness of response selection of a first task. This influence demonstrates simultaneity of semantic retrieval and response selection. It thus seems that the bottleneck account holds for combinations of tasks sharing the same general type of processes at a rather abstract level. That is, while the present experiments require retrieval from two types of long-term memory components (i.e., semantic and episodic) resulting in a retrieval bottleneck, choice RT tasks require activations of cue-response mapping rules from working memory (cf. Thomson et al., 2010). Thus, we suggest that the present findings in combinations with previous ones support the idea of long-term memory and working memory as being autonomous components (Baddeley, 2000, 2003). Equivalent bottleneck mechanisms might operate within each component but not across components.

Concluding remarks

We investigated how dual memory-retrieval takes place when one retrieval is from newly established episodic memory and one is from long established semantic memory. The findings support the existence of a retrieval bottleneck and learned parallelism through cue-level response chunking, just has been demonstrated for the case of dual episodic retrieval (Nino & Rickard, 2003; Strobach et al., 2014). These results are consistent with a global memory bottleneck account instead of a memory system specific account. Those results occurred both when only new cues were included in the transfer phase (Experiment 1), and when both old and new cues were randomly intermixed in the transfer phase (Experiments 2 and 3). We were further able to support a cue-level chunking account of learned parallelism with English antonym pairs in English native speakers (Experiment 1) and with German antonym pairs in German native speakers (Experiments 2 and 3) as well as with a moderate (Experiments 1 and 2) and an extensive amount of practice (Experiment 3), although in the latter case there was also some evidence of a task-level parallel retrieval mechanism. These results extend the prior knowledge on dual memory retrieval (e.g., Strobach et al., 2014) to a new combination of memory retrievals, memory systems, as well as across different languages and across different practice levels.

Acknowledgements

We would like to thank Miwa Yamazaki, Sabine Kaspers, Kristina Schindler, and Fatma Ates for their assistance with data collection. The authors take responsibility for the integrity of the data, the accuracy of the data analyses, and have made every effort to avoid inflating statistically significant results.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This study was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) under Grant number

STR 1223/1. Furthermore, it was supported by a grant of the German Academic Exchange Service (Deutscher Akademischer Austauschdienst, DAAD) to T.S. (second author).

ORCID iD

Franziska Orscheschek https://orcid.org/0000-0001-6211-3105

References

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

Baddeley, A. (2000). Short-term and working memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 77–92). New York, NY: Oxford University Press.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839.

Bowen, H. J., & Kensinger, E. A. (2017). Cash or credit? Compensation in psychology studies: Motivation matters. *Collabra: Psychology*, *3*, 12.

Brase, G. L. (2009). How different types of participants payments alter task performance. *Judgment and Decision Making*, *4*, 419–428.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*, 51–57.

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*, 1058–1079.

Fischer, R., Miller, J., & Schubert, T. (2007). Evidence for parallel semantic memory retrieval in dual tasks. *Memory & Cognition*, *35*, 1685–1699.

Fischer, R., & Plessow, F. (2015). Efficient multitasking: Parallel versus serial processing of multiple tasks. *Frontiers in Psychology*, *6*, 1–11.

Friederici, A. D., & Chomsky, N. (2017). *Language in our brain: The origins of a uniquely human capacity*. Cambridge: The MIT Press.

Gaigg, S. B., Bowler, D. M., & Gardiner, J. M. (2014). Episodic but not semantic order memory difficulties in autism spectrum disorder: Evidence from the Historical Figures Task. *Memory*, *22*, 669–678.

Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition*, *26*, 263–276.

García-Lázaro, H. G., Ramirez-Carmona, R., Lara-Romero, R., & Roldan-Valadez, E. (2012). Neuroanatomy of episodic and semantic memory in humans: A brief review of neuroimaging studies. *Neurology India*, *60*, 613–617.

Gong, Y., Ericsson, K. A., & Moxley, J. H. (2015). Recall of briefly presented chess positions and its relation to chess skill. *PLoS ONE*, *10*, e0118756.

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*, 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.

Hermann, D. J., Conti, G., Peters, D., Robbins, P. H., & Chaffin, R. J. (1979). Comprehension of antonymy and the generality of categorization models. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 585–597.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*, 785–813.

Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C. S., ... Buchanan, E. (2013). The semantic priming project. *Behavior Research Methods*, *45*, 1099–1114.

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

JASP Team. (2017). JASP (Version 0.8.1.1) [Computer software].

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*, 573–603.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

Logan, G. D., & Schulkind, M. D. (2000). Parallel memory retrieval in dual-task situations: I. Semantic memory. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1072–1090.

Loureiro, I. S., & Lefebvre, L. (2016). Distinct progression of the deterioration of thematic and taxonomic links in natural and manufactured objects in Alzheimer's disease. *Neuropsychologia*, *91*, 426–434.

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, *7*, 618–630.

Maquestiaux, F., Laguë-Beauvais, M., Ruthruff, E., & Bherer, L. (2008). Bypassing the central bottleneck after single-task practice in the psychological refractory period paradigm: Evidence for task automatization and greedy resource recruitment. *Memory & Cognition*, *36*, 1262–1282.

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

Nelson, D. L., Dyrdal, G. M., & Goodmon, L. B. (2005). What is preexisting strength? Predicting free association probabilities, similarity ratings, and cued recall probabilities. *Psychonomic Bulletin & Review*, *12*, 711–719.

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*, 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*, 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.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 411–421.

Orscheschek, F., Rickard, T., Schubert, T., & Strobach, T. (2018). *Dual-memory retrieval efficiency after practice: Effects of strategy manipulations*. Manuscript in preparation.

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: The MIT Press.

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research*, *66*, 180–194.

Rajah, M. N., & McIntosh, A. R. (2005). Overlap in the functional neural systems involved in semantic and episodic memory retrieval. *Journal of Cognitive Neuroscience*, *17*, 470–482.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 435–451.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., ... Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*, 359–379.

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., 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.

Schmolck, H., Kensinger, E. A., Corkin, S., & Squire, L. R. (2002). Semantic knowledge in patient HM and other patients with bilateral medial and lateral temporal lobe lesions. *Hippocampus*, *12*, 520–533.

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 Bioscience*, *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*, 101–108.

Simmons, S., & Estes, Z. (2006). Using latent semantic analysis to estimate similarity, *Proceedings of the 28th Cognitive Science Society* (pp. 2169–2173). Austin, TX: Cognitive Science Society.

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

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

Strobach, T., Frensch, P. A., Müller, H., & Schubert, T. (2015). Evidence for the acquisition of dual-task coordination skills in older adults. *Acta Psychologica*, *160*, 104–116.

Strobach, T., & Schubert, T. (2017). No evidence for task automatization after dual-task training in younger and older adults. *Psychology and Aging*, *32*, 28–41.

Strobach, T., Schubert, T., Pashler, H., & Rickard, T. (2014). The specificity of learned parallelism in dual-memory retrieval. *Memory & Cognition*, *42*, 552–569.

Thomson, S. J., Watter, S., & Finkelshtein, A. (2010). Parallel response selection in dual-task situations via automatic category-to-response translation. *Attention, Perception, & Psychophysics*, *72*, 1791–1802.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*, 1–25.

Tulving, E., & Hastie, R. (1972). Inhibition effects of intralist repetition in free recall. *Journal of Experimental Psychology*, *92*, 297–304.