Author-final version accepted for publication in the Journal of Experimental Psychology:General

Storage and processing in working memory: A single, domain-general resource explains multi-tasking

Evie Vergauwe

Faculty of Psychology and Educational Sciences, University of Geneva

Claudia C. von Bastian

Department of Psychology, University of Sheffield

Ralitsa Kostova and Candice C. Morey

School of Psychology, Cardiff University

Word count: 9'425 (excluding title, references, author affiliations, acknowledgements, figure, and figure legends, but including abstract)

Author note. Address correspondence to either Dr. Evie Vergauwe (evie.vergauwe@unige.ch) or Dr. Candice C. Morey (MoreyC@cardiff.ac.uk). Materials associated with this paper are available at osf.io/jpz3u. The data and ideas in the manuscript were presented at the 2020 Meeting of the Experimental Psychology Society and the 2020 Annual meeting of the Psychonomic Society (virtual meetings, due to covid-19). This work was completed with support from British Academy/Leverhulme Small Research Grant SRG18R1\180622 to Candice Morey and Swiss National Science Foundation Grant PCEFP1_181141 to Evie Vergauwe

Abstract

An ongoing major debate centers around whether multi-tasking in working memory, that is, performing several mental activities at once, is supported by multiple specialized *domain-specific* or by a single-purpose *domain-general* cognitive resources. Working memory theories differ in their explanations and predictions about when performing two mental tasks causes performance failures, versus when two processes can be carried out concurrently with negligible cognitive costs. In particular, the predictions of domain-specific and domain-general views on working memory are in conflict with one another when it comes to the cognitive cost associated with concurrent verbal and visuo-spatial processing and storage tasks. Previous tests of these predictions using traditional methods have led to ambiguous and inconsistent conclusions, however. To make critical progress in this theoretical debate, we used a radically different approach combining Bayesian state-trace analysis with an experimental design fully crossing processing and storage tasks differing only in the domain of representation (verbal vs. visuo-spatial). Across two experiments, we show unambiguously that a single, domain-general factor can account for briefly maintaining verbal and visuo-spatial information in a multi-tasking scenario.

Storage and processing in working memory: A single, domain-general resource explains multi-tasking

Multi-tasking scenarios involving remembering while doing something else are basic human skills ubiquitous in everyday life. For example, when we pay cash, we must remember how much we already gave while calculating how much more is needed; when we take notes in class, we need to maintain and write the beginning of the sentence while listening further; when we cook a new dish, we need to remember the list of ingredients while we search for the eggs in the fridge, and when we drive, we frequently do so while holding a conversation. Understanding how information is maintained under such circumstances is vital for applying psychological principles outside the lab. Fundamentally, understanding the nature of the limited resources supporting dual-task performance is a critically important step towards a more comprehensive understanding of the limits of human information processing. At the practical level, understanding the nature of limited cognitive resources is crucial for understanding performance in real-world multi-tasking situations, which are increasingly common, and for understanding the disrupted dual-task performance typically observed in neurodegenerative disorders such as Alzheimer's disease (Foley et al., 2011). The nature of the limited resources supporting mental activities is thus a central issue affecting a wide range of scientists and practitioners.

An ongoing debate centers around whether mental processes are supported by multiple specialized systems (the modular view on cognitive resources; Coltheart, 1999; Fodor, 1983; Navon & Gopher, 1979) or by a single general-purpose system (the unitary view on cognitive resources; Kahneman, 1973; Norman & Bobrow, 1975). One of the main characteristics of a modular view on cognition is that mental activities are assumed to be supported by domain-specific resources, resulting in no (or a very limited) cognitive cost when performing two mental tasks that pertain to different domains. In particular, separate resources have been proposed for verbal mental activities (such as talking, reading, remembering numbers) and for visuo-spatial mental activities (such as wayfinding, processing colors, remembering locations; Wickens, 1980) . For example, Wickens' (1984) multiple resource theory of multi-tasking behavior proposes separate pools of resources for verbal and spatial processing codes. Similar domain-specific assumptions about the nature of limited cognitive resources were put forward by the immensely popular multiple-component model of working memory (Baddeley & Hitch, 1974; Baddeley & Logie, 1999). *Working memory* refers to the limited-capacity cognitive system responsible for concurrent processing and temporary memory[1]. Working memory theory attempts to explain and predict when performing two mental tasks caused performance failures, versus when two processes could be carried out concurrently with negligible cognitive costs (Morey, 2018). Different working memory models propose different collections of resources that predict quite different multi-tasking outcomes.

Different working memory models predict different multi-tasking outcomes

Within the multiple-component model, it is proposed that working memory consists of multiple resources, many of which are domain-specific. A distinction is made specifically between resources supporting verbal memory and resources supporting visuo-spatial memory (Baddeley, 1986; Baddeley & Logie, 1999; Baddeley et al., 2019; Berry et al., 2019; Shah & Miyake, 1996). The model proposes two entirely separate subsystems for holding verbal and visuo-spatial information, the phonological loop and the visuo-spatial sketchpad, respectively. These subsystems are assumed to function independently from each other, such that increasing demands on one subsystem do not affect ongoing activities of the other subsystem. According to this view, the nature of the limited resources underpinning working memory, and thus the nature of the limited resources supporting dual-task performance, is inherently domain-specific. As a result, whereas some interference between two verbal or between two visuo-spatial activities might occur, domain-specific views argue that this interference cannot arise because of conflicts between holding verbal and visuo-spatial information. Because maintenance occurs in parallel in these separate short-term stores, small amounts of verbal and visuo-spatial information are always protected from general dual-task interference. Patterns showing substantial interference with memoranda from the same domain and negligible interference with information from different domains, referred to as *selective interference*, are predicted by domain-specific views on working memory, and bolster the idea that separate, domain-specific modules support ongoing cognition.

Other models of working memory have proposed other collections of resources to support concurrent processing and storage. In many of these models, a domain-general pool of resources is proposed, which supports both verbal and visuo-spatial activities (Barrouillet et al., 2004; Cowan, 1995; Engle et al., 1999). For example, several models propose that both processing and storage activities are supported by domain-general attention (Barrouillet et al., 2004; Cowan, 1995; Engle et al., 1999; Oberauer, 2013), which must be shared between verbal and visuo-spatial activities when these are performed concurrently. As such, according to domain-general views, the nature of the limited resources supporting dual-task performance is primarily domain-general, and substantial levels of interference between concurrent tasks are to be expected, regardless of the domain involved. As a result, domain-general views of working memory predict non-negligible, *general interference* between concurrent verbal and visuo-spatial activities.

Clearly, the predictions of domain-specific and domain-general views are contradictory when it comes to the cognitive cost associated with concurrent verbal and visuo-spatial activities. Theoretically, the predictions of domain-specific and domain-general views can be tested straightforwardly by creating dual-task situations using a variety of task combinations to assess whether remembering verbal versus visuo-spatial information is differently impacted by engaging in concurrent verbal versus visuo-spatial processing. The daily-life problem of having to store information while concurrently processing other incoming information has been approximated in the laboratory by using complex span tasks. Complex span tasks (Conway et al., 2005) require participants to remember short lists (e.g., a list of words) while carrying out a processing task in between each memory item (e.g., solving an arithmetic problem between presentations of each word). In this task, it is possible to combine verbal memory lists with either verbal or visuo-spatial processing tasks, in the same way as it is possible to combine visuo-spatial memory lists (e.g., a list of spatial locations on screen) with either verbal or visuo-spatial processing tasks. Theoretically, comparing the levels of performance between these dual-task situations, and especially between same-domain (i.e., verbal memory plus verbal processing, or visuo-spatial memory plus visuo-spatial processing) and different-domain task combinations (i.e., verbal memory plus visuo-spatial processing, or visuo-spatial memory plus verbal processing) should disentangle domain-specific and domain-general views on working memory; domain-general views predict interference effects in both same-domain and different-domain task combinations, whereas domain-specific views anticipate more pronounced interference effects in same-domain combinations than in different-domain combinations (but see Cocchini et al., 2002 and Duff & Logie, 2001 for a version of the multiple-component model that does not necessarily anticipate interference between memory and processing).

Studying multi-tasking outcomes has not led to definitive conclusions

Several studies have followed this approach, using complex span tasks in which verbal and visuo-spatial memory tasks are combined with verbal and visuo-spatial processing tasks (Bayliss et al., 2003; Shah & Miyake, 1996; Thalmann & Oberauer, 2017; Vergauwe et al., 2010, 2012). However, this approach has not led to definitive conclusions; in fact, despite more than 20 years of investigation, the same basic debate continues. We argue that there are two main reasons for this lack of resolution and propose a novel approach addressing both.

First, some proponents of both domain-specific and domain-general views have obtained similar patterns of results but reached opposite conclusions. For example, Shah and Miyake (1996) combined verbal (series of words) vs. visuo-spatial (series of arrows) memory with verbal (sentence verification) vs. visuo-spatial (mental rotation) processing, obtaining data from four different complex span tasks. An asymmetry was revealed between verbal and visuo-spatial memory, indicating that verbal memory might be more sensitive to the nature of concurrent same-domain processing than visuo-spatial memory. Visuo-spatial memory was affected by both visuo-spatial and verbal processing, especially in conditions with a higher number of to-be-remembered items. Still, the overall weaker memory performance in same-domain combinations relative to cross-domain combinations led Shah and Miyake to conclude that their results provided evidence for separable resources for verbal and visuo-spatial mental activities in working memory. A similar pattern of interference was reported by Vergauwe, Barrouillet, and Camos (2010), who also combined verbal (series of letters) vs. visuo-spatial (series of locations) memory with verbal (semantic judgment) vs. visuo-spatial (spatial fit judgment) processing. They also found an asymmetric pattern: verbal recall accuracy was particularly affected by verbal relative to visuo-spatial processing, whereas visuo-spatial recall accuracy was equally affected by both verbal and visuo-spatial processing. However, because an effect of cognitive load was found in all four complex span tasks (i.e., poorer recall as attentional demands of processing task increased), Vergauwe et al. concluded that domain-general resources must be shared between verbal and visuo-spatial activities in working memory – a conclusion opposite to that drawn by Shah and Miyake. The possibility to flexibly interpret the same patterns stymies convergence on a common interpretation, which is vital for theoretical progress.

Second, examining interference patterns in different task combinations across the verbal and visuo-spatial cognitive domains is further complicated by factors that are difficult to control between tasks. For example, people tend to encode visuo-spatial stimuli verbally and, so, may suffer from verbal distraction even in visuo-spatial memory conditions (Shah & Miyake, 1996). Although it is more difficult to imagine the reverse, verbal stimuli may also be encoded so that their visual characteristics are represented (Logie et al., 2000), which logically poses the reverse problem. Moreover, in manipulating the task domain, more than only the form of the mental representation differs in many studies. Response modalities, the timing of stimulus presentations, the speed with which recall responses may be made, all might vary along with the presumed domain of the mental representation (Thalmann & Oberauer, 2017). For example, in addition to differing in the nature of the mental representations involved, verbal and visuo-spatial processing often differ in the amount of information to be processed in each processing step (reading an entire sentence vs. judging a single image; Shah & Miyake, 1996), or processing one word vs. a spatial configuration of several elements (Vergauwe et al., 2010). Similarly, verbal and visuo-spatial memory often differ in the nature of recall response (oral recall vs. mouse clicking; Shah & Miyake, 1996), or oral recall vs. pointing; Bayliss et al., 2003). Confounds between the domain of the involved representations and differences between other task properties hinder straightforward interpretation of interference patterns in terms of domain-general vs. domain-specific stores. As a result, after decades of work, there is still disagreement about whether working memory models must propose one or multiple short-term stores.

Traditional methods for examining patterns of selective interference between concurrent activities in dual-task research designs have not provided a definitive solution. Though widely considered the gold standard for discovering functionally specific modules, the dissociations frequently reported in dual-task working memory research do not provide unambiguous evidence that multiple processes underlie performance (Dunn & Kirsner, 1988). Thus, we need a novel method capable of unambiguously establishing whether remembering verbal vs. visuo-spatial information is differentially impacted by engaging in verbal vs. visuo-spatial processing.

The current study: A novel approach to a longstanding debate

This need can be satisfied by applying Bayesian state-trace methodology (Cox & Kalish, 2019; Prince et al., 2012) to data from complex-span tasks, which allows for addressing the first reason for a lack of a definitive solution (i.e., the possibility to flexibly interpret the same patterns of interference). State-trace analysis is a method specifically designed to test the dimensionality of psychological constructs (see Stephens et al. (2019) for a recent paper on the use and strength of state-trace analysis to address theoretical questions in terms of single vs. multiple underlying systems in the field of experimental psychology). It requires implementing a particular experimental design that has not been carried out before, comprising three elements: state, dimension, and trace (see Figure 1). The *state* variable allows for differentiating between two hypothetical latent resources, in this case, verbal and visuo-spatial short-term memory buffers tapped by testing memory for verbal and visuo-spatial information. The *dimension* variable may disrupt the workings of the hypothesized latent constructs. Here, this is accomplished by manipulating the domain of the processing task (i.e., verbal or visuo-spatial). Importantly, we carefully manipulated the domains of storage and processing so that only the representational domain differed between verbal and visuo-spatial task versions, avoiding potential confounds arising from also varying the complexity of the stimuli to be processed, their timings, or the mode of responding, and thereby addressing the second reason for a lack of a definitive solution (i.e., confounds between the domain of the involved representations and differences between other task properties). The *trace* element of the design affords a view of how the latent resources are affected at different levels of performance. To ensure that for each participant and memory condition, we observed recall performance that systematically decreased from near-ceiling, we manipulated the number of memoranda (i.e., memory load) per sequence. To create sufficient scope to detect diverging traces, it is important that for each participant, the levels of the trace variable produce wide ranges of performance. Prince et al. advise that the levels of the trace variable may differ per participant or per other levels in the design to ensure adequate conditions for interpreting the outcome.

State-trace methodology is fundamentally graphical (see Figure 1). Performance on each level of the state factor is plotted separately for each level of the dimension factor. This results in two traces, one for each level of the dimension factor (i.e., verbal vs. visuo-spatial processing, represented in Figure 1 by the differently-colored lines), with one point per line for each level of the trace variable (i.e., level of load). When the task is designed so that individual participants perform in comparable ranges spread downward from off-ceiling on both memory tasks, two possible outcomes can occur, shown in Figures 1a and 1b. If the dimension factor (verbal vs. visuo-spatial processing) affects one level of the state task (verbal vs. visuo-spatial memory) more than another, the traces diverge (see upper panel Figure 1). Unlike the ambiguous interactions previously observed when contrasting same- and cross-domain combinations in complex span tasks, observing this *multidimensional* pattern in the state-trace analysis cannot be explained by a hypothetical system with only one latent factor supporting both levels of the state task. Such a pattern must be interpreted as arising from a multidimensional system which is differentially impacted by the two levels of the dimension factor. However, if the dimension factor affects both levels of the state task similarly, then a monotonic pattern like the one in Figure 1b occurs, whereby data of the state tasks fall on a single, monotonically increasing curve in the state-trace plot. This *unidimensional* pattern can be explained by positing a single latent factor. Observing a pattern like this would mean that a simpler, single-factor model might explain the pattern, see lower panel Figure 1. Applying Bayesian state-trace methodology solves the problem previously inherent in accepting the unidimensional solution, namely that one must be persuaded of the absence of evidence for multidimensionality[2].

In particular, Cox and Kalish's (2019) Bayesian method[3] for computing the degree to which a dataset yields a monotonic state-trace provides an outcome statistic, $\widehat{M}$, which quantifies the relative probability that two outcome variables are jointly monotonic by comparing the posterior probability of monotonicity against a prior probability of monotonicity. $\widehat{M}$ provides evidence capable of supporting either a uni- or a multi-dimensional pattern by assessing whether the derivatives arising from the latent covariance structure of the average effects of conditions jointly on each outcome variable share the same sign. $\widehat{M}$ is calculated by estimating the proportion of the distribution of joint derivatives sharing the same sign, smoothing this estimate to better reflect uncertainty in the true posterior distribution (denoted $\tilde{p}_M$) and dividing this by the sum of $\tilde{p}_M$ and the chosen prior probability of monotonicity. In our case, we used an uninformed prior to reflect ambiguity about how previous evidence should be interpreted. $\widehat{M}$ may then be interpreted as the degree to which the data changed our expectation about the joint monotonicity of the outcome measures: this is what makes it Bayesian. Combined with an experimental design that focuses on contrasting verbal and visuo-spatial storage, this methodology provides the evidence needed to declare whether multiple domain-specific short-term stores underlie complex span performance or not.

It should be noted that, like the traditional methods which examine patterns of selective interference between concurrent activities in dual-task designs, our state-trace approach works under the assumption that processing and storage activities can interfere with each other in working memory. While most domain-general and domain-specific views agree on this point (e.g., Barrouillet et al., 2004; Cowan, 1995; Engle et al., 1999; Oberauer, 2013 for domain-general views and Baddeley, 2000; Baddeley &

Hitch, 1974; Baddeley & Logie, 1999; Baddeley et al., 2019; Repovs & Baddeley, 2006; Shah & Miyake, 1996 for domain-specific views), there is at least one version of the multiple-component model that does not necessarily adhere to this assumption (see Cocchini et al., 2002; Duff & Logie, 2001). As a result, this particular version of the multiple-component model cannot be tested using our state-trace approach. However, to foreshadow, consistent with the notion that processing and storage can interfere with each other, and in line with a large body of research, our results showed clear dual-task costs when we compared dual-task performance with single-task performance. Moreover, consistent with the notion that processing and storage can interfere with each other in working memory, multiple previous studies have shown that memory performance is a direct function of the demands of concurrent processing (Barrouillet et al., 2004, 2007, 2011), and that memory performance depends on the amount of interference in working memory between memory items and representations involved in concurrent processing (Oberauer, Farrell, et al., 2012; Oberauer, Lewandowsky, et al., 2012; Oberauer & Lange, 2008). Together, these findings rule out any version of the multiple-component model that does not predict interference between processing and storage, and it is in line with our assumption that processing and storage can interfere with each other when performed concurrently.

In sum, the goal of this study was to test whether a single, domain-general factor can account for maintenance in working memory, or alternatively, whether we need to propose a model that includes multiple domain-specific factors. To that end, we applied a Bayesian state-trace approach using complex span tasks that carefully manipulated the domain of concurrent processing and storage activities, while relying on widely accepted assumptions and commonly-used tasks. In two experiments, participants completed complex span tasks with verbal (aurally-presented non-words) and visuo-spatial (visually-presented spatial locations) memoranda, fully crossed with verbal (rhyme judgment) and visuo-spatial (symmetry judgment) processing tasks, yielding four task combinations. Memory items were presented sequentially, with presentation durations and serial reconstruction procedure matched across verbal and visuo-spatial memory tasks. We implemented an adaptive algorithm that automatically adjusted list lengths to ensure that accuracy ranged from near-ceiling to clearly off-ceiling for each participant (see Supplementary materials 2 for a detailed description). In Experiment 1, we started with set sizes 2, 3, and 4, whereas in Experiment 2, we started with set sizes 2, 4, and 6. Processing tasks were modeled after those described by Jarrold et al. (2011) and selected for yielding equivalent accuracy rates and response times (which was confirmed by our data, see below). Both processing tasks consisted of two simultaneously presented items to which participants responded via mouse click; verbal processing consisted of judging whether two letters rhymed, whereas visuo-spatial processing consisted of judging whether two inverted letters shared an axis of symmetry. The novelty of our approach lies in (1) the specific combination of tasks, (2) the explicit effort to isolate only the domain of the cognitive representations involved in the different tasks, and (3) the use of Bayesian state-trace analysis to provide a strict and straightforward test of domain-specific vs. domain-general accounts of working memory.

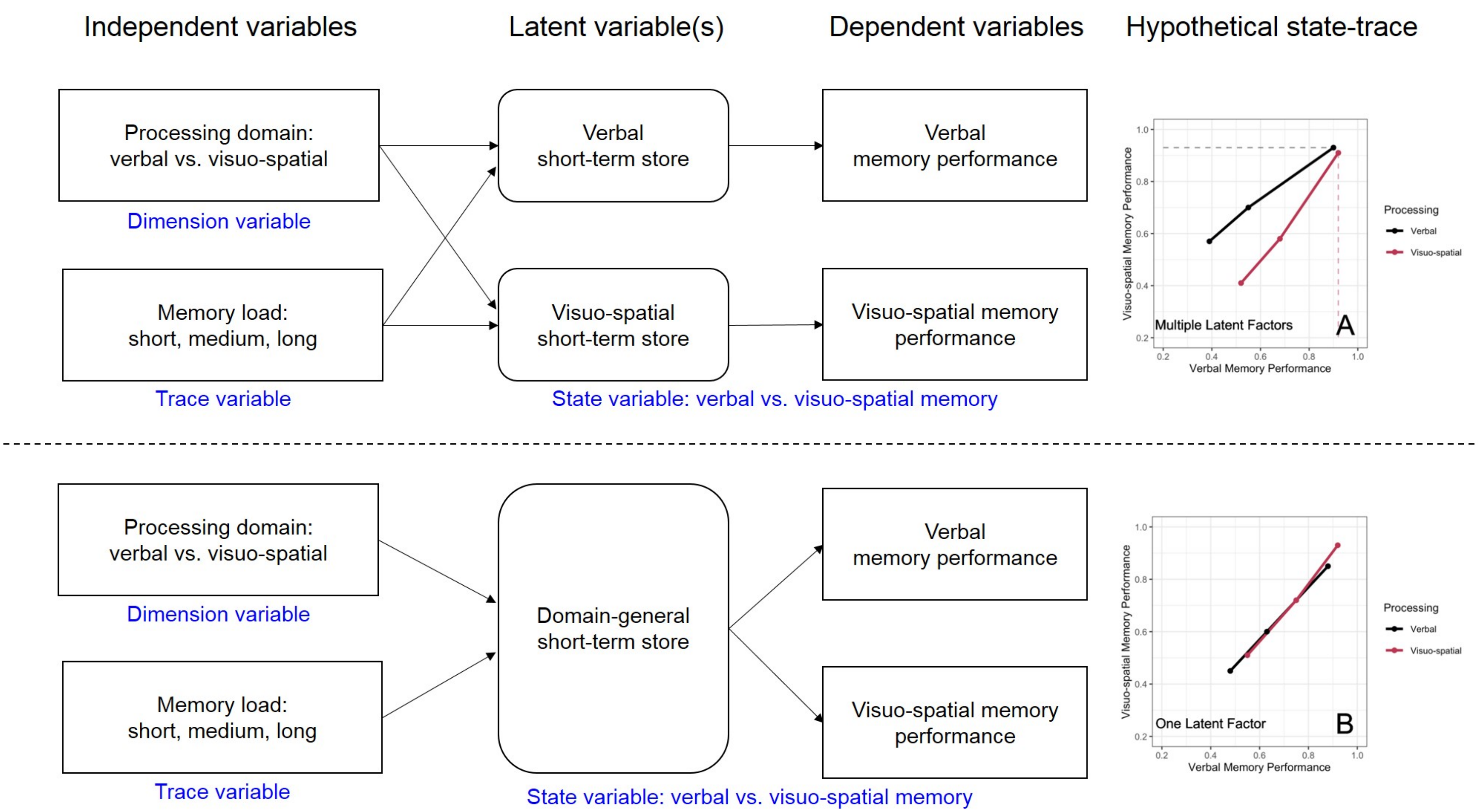

Figure 1. Schematic illustration of multidimensional (upper panel) and unidimensional (lower panel) accounts of working memory. **The upper panel** shows a multidimensional account, proposing two separate, domain-specific stores (one for verbal information, and one for visuo-spatial information; implemented in our study by using verbal vs. visuo-spatial memory tasks: memory for nonwords vs. memory for locations, respectively), that are both affected by memory load (implemented in our study by using memory lists of varying length) and differently affected by the domain involved in a concurrent processing task (implemented in our study by using verbal vs. visuo-spatial processing tasks: rhyme vs.

symmetry judgment). If multiple, domain-specific latent factors underlie working memory storage, diverging hypothetical memory performance traces as shown in (A) should be observed, when plotting verbal memory performance (memory accuracy for nonwords, in our study) and visuo-spatial memory performance (memory accuracy for locations, in our study) against each other, as a function of Processing domain (dark vs. magenta line) and Memory load (different points in the graph). We do not expect performance on either task to sit near these dashed lines; their presence helps to show that many possible patterns (including asymmetric ones) short of complete absence of interference could reflect the operation of multiple latent factors. **The lower panel** shows a unidimensional account, proposing a single, domain-general store for both verbal and visuo-spatial information), that is affected by memory load and domain involved in a concurrent processing task. If a single, latent factor supports working memory storage, a monotonic pattern as shown in (B) should be observed when plotting verbal memory performance and visuo-spatial memory performance against each other, as a function of Processing domain (dark vs. magenta line) and Memory load (different points in graph).

Experiment 1

Method

Participants

We recruited 21 healthy adults from the Cardiff area[3]. Participants received £32 for completing 3 sessions lasting approximately 90 minutes each. Sessions were spaced by at least 24 hours, and at most 1 week. We checked for a minimal level of performance in the first session before inviting participants to take part in subsequent sessions. To qualify to continue after the first session, participants must have achieved at least 80% accuracy on each of the processing tasks. We applied the same criterion to each participant's complete data set for inclusion in the eventual analysis. One participant was prohibited from completing the study due to failure to meet our pre-specified performance target in the first session. Another participant was excluded after data collection because the average accuracy on the rhyme judgment task across all sessions fell below 80%. The final analyzed sample therefore included 19 adults (17 female) between the ages of 19 and 26 (M= 22.58, SD = 1.95). All participants reported normal or corrected-to-normal vision, normal hearing, spoke fluent English, and presented with no history of neurological dysfunction that required the use of medication likely to affect cognitive abilities. One participant completed only two sessions due to inability to negotiate a time for the final session with the researcher. The study protocol was approved by the Ethics Committee of Cardiff University.

Materials

Data were collected on a personal computer located in a private sound-attenuated testing booth. Figure 2 illustrates the general task procedure. Verbal memoranda were aurally-presented nonwords selected randomly from sets of 30 non-words chosen using the English Lexicon Project (Balota et al., 2007). We queried the database for 2-syllable nonwords between 5 and 8 letters long with no orthographic neighbors. We reduced the resulting candidate set of nonwords to 30, choosing pronounceable nonwords that we judged most dissimilar to legitimate words, and whose pronunciation was executed correctly by the text-to-speech system built into Mac OS. Each .wav file was created using the British-accented artificial voice *Kate* available in Mac OS. These nonwords ranged in pronunciation time from 0.38-0.59 seconds. We divided the non-words pseudo-randomly into three lists of ten. Lists were set so that each list included items with the same initial and ending sounds, to discourage strategies of selectively attending part of the nonword (see Supplementary materials 1). One set was randomly selected per participant and session such that each participant experienced each list of nonwords once across the three sessions, to prevent participants carrying over familiarity with the nonwords across sessions. We used non-words rather than words or letters as verbal memory material, based on the expected range of performance which we wanted to match closely that of visuo-spatial memory material.

Visuo-spatial memoranda were chosen pseudo-randomly from an invisible 8-by-8 grid in the central region of the monitor (approx. 13.6 x 13.6 cm). For each session, and for each participant, a set of 10 positions were selected such that, for any given participant, a new set of 10 positions was used in each session. Each position was marked with a square and appeared highlighted in blue onscreen for 380-590 ms, matching the variability in the pronounced durations of the nonwords (the exact durations of the positions were randomly chosen out of a list containing the exact durations of the non-words). Any two positions in the set were at least one grid cell (approx. 1.7 cm) apart. Using comparable verbal and visuo-spatial stimuli, Morey and

Miron (2016) observed similar performance between verbal and visuo-spatial serial order reconstruction tasks for 5-item sequences.

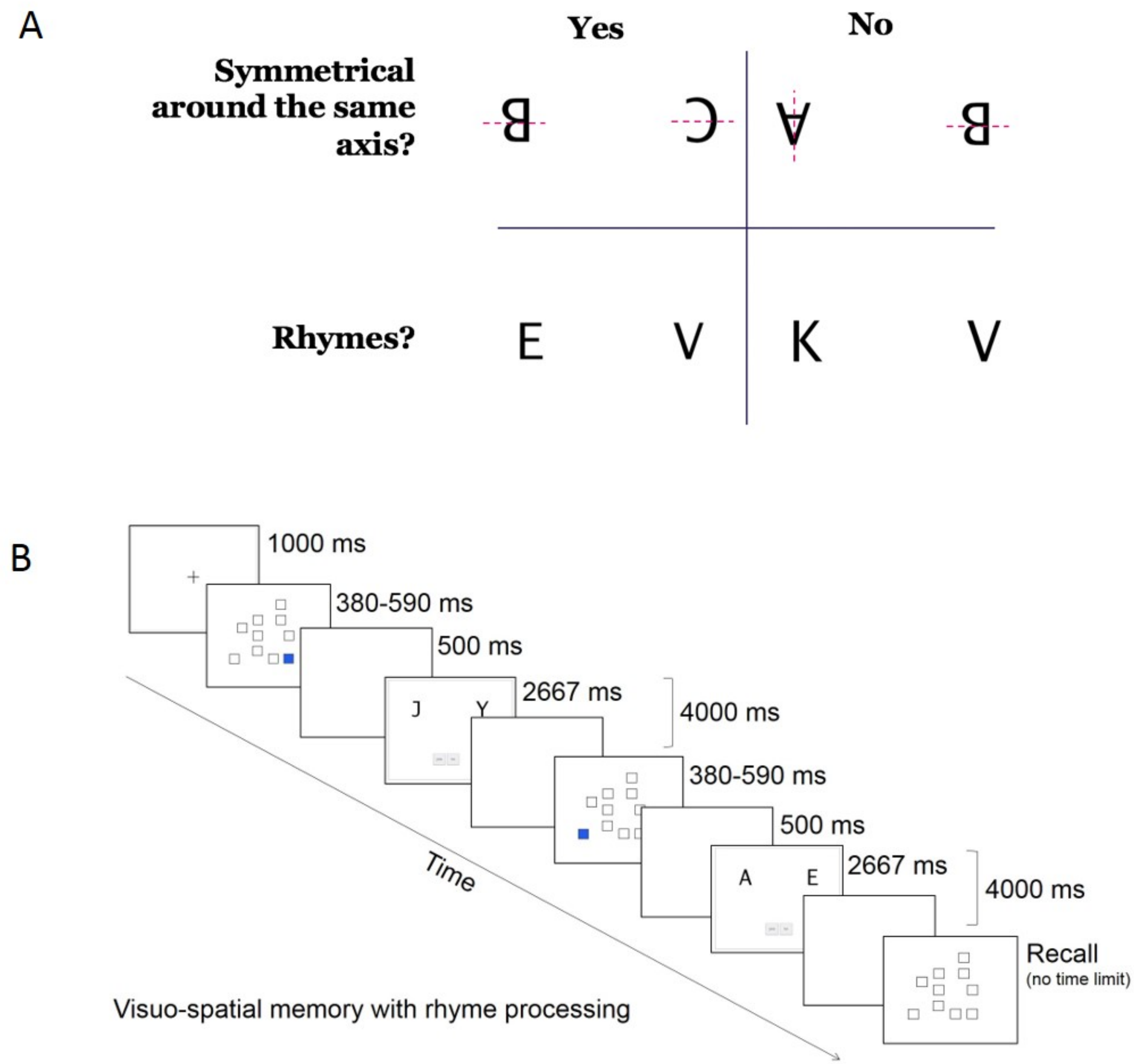

Figure 2. Panel A: Two processing tasks requiring decisions based on visuo-spatial or phonological aspects of the stimuli. In the symmetry task, participants judged whether two inverted, mirrored letters shared the same axis of symmetry (dotted lines are shown to denote each letter's axis of symmetry; these were not shown during the task). In the rhyme task, participants judged whether the pronunciation of upright letters rhymed. Examples of true and false pairings are provided for both tasks. Panel B: Events in a trial procedure combining visuo-spatial memoranda with rhyme processing judgments. In this example, both rhyme stimuli should elicit "No" responses. Participants completed blocks during each session fully crossing verbal and visuo-spatial memoranda with verbal and visuo-spatial processing.

We used processing tasks comparable to those described by Jarrold, et al. (2010, 2011), which require participants to perform different judgments on identical stimuli. Figure 2 depicts examples of the processing stimuli. Participants were always shown two uppercase letters presented side-by-side in 96-point Lucida Sans Typewriter font in the center of the screen, spaced by 9 cm. In the verbal conditions, participants judged whether the two letters rhymed (e.g., C and E rhyme; J and T do not). In the visuo-spatial condition, letters were presented upside-down and mirrored to discourage participants from reading them. Participants judged whether the letters shared an axis of symmetry (e.g., A and T are both symmetric around the vertical axis, while C and V are symmetrical around the horizontal and vertical axes respectively). Letters were drawn from restricted 10-letter sets (rhyme judgements: A, C, D, E, I, J, K, T, V, Y; symmetry judgments: A, B, C, D, E, K, M, T, V, Y), such that pairs eliciting "yes" and "no" responses occurred on 50% of trials in both versions of the task.

Given that previous research has shown that sustained practice of complex span tasks substantially increases the number of items typically recalled (von Bastian & Oberauer, 2013), we adjusted the storage list lengths based on individual participant performance to ensure that accuracy ranged from near-ceiling to clearly off-ceiling for each participant. Specifically, at the end of each session, we computed the proportion of correctly recalled items per complex span task and per participant (independently of performance in the processing task). Based on previous work with similar stimuli, we expected that, without sustained practice, list lengths of 2, 3, and 4 would elicit performance in the appropriate ranges; young adults typically remember about 2.5 to 3 items in dual-task situations requiring maintenance of series of locations or series of non-words (Camos et al., in press; Morey & Miron, 2016; Vergauwe et al., 2010, 2012).

Thus, all participants began with list lengths 2, 3, and 4 in session 1. If the individual accuracy score was 80% or greater, the adaptive algorithm automatically increased list lengths by 1 item in the following session (e.g., list lengths 3, 4, and 5, instead of list lengths 2, 3, and 4; see Supplementary materials 2 for a detailed description of the adaptive algorithm). If accuracy was below 80%, the list lengths remained the same. As the adaptive algorithm was implemented in the experiment presentation software, accuracy calculations and adjustment decisions occurred in an experimenter-blind manner.

Stimulus presentation and response collection were controlled via custom programs produced with Tatool Web, an open-source JavaScript-based platform for psychological research (www.tatool-web.com; von Bastian et al., 2013). In all tasks, participants responded to the processing tasks via mouse click and entered their mnemonic response by choosing from the options in each 10-item set via mouse. These open-source experimental tasks are publicly available on Tatool Web (linked via the experimental protocol on our Open Science Framework page, https://osf.io/jpz3u/).

Procedure

Most participants in the final sample completed approximately 5 hours of testing spread across three separate sessions, yielding 576 complex span trial administrations (16 trials x 3 set sizes x 4 complex span tasks x 3 sessions) per participant in addition to practice trials. One participant completed only two sessions due to scheduling difficulties.

In their first session, participants who provided written consent were introduced to each storage and processing task separately. For each storage-and-processing combination, participants practiced (1) the processing task only, (2) the storage task only, and (3) the storage-and-processing tasks interleaved (see Redick et al., 2012) for a similar approach). Within each practice block, participants were first trained on the processing task. Participants completed 100 letter pairs successively presented on screen. Each letter pair remained on screen until a response was detected, and the next letter pair was presented after a 500-ms blank screen. Visual feedback (green thumbs-up or red thumbs-down) was presented during the presentation of the blank screen. At the end of 100 letter pairs, participants' accuracy across the last 20 judgments was displayed on screen. Participants who did not achieve an accuracy rate of at least 80% on the last 20 judgments were asked to perform another set of 100 judgments. Within each block, participants were then trained on the storage task. After a brief explanation of a single trial, participants were presented with six practice trials in which no processing items were presented in between the memory items. At the end of each trial, participants received visual feedback on their memory performance (green thumbs-up or red thumbs-down). Finally, participants were given practice with the combined storage and processing tasks. They completed four series of two to-be-remembered items, interleaved with processing judgements. Participants experienced four such practice blocks in random order, one for each of the four storage-and-processing conditions (verbal storage - verbal processing, verbal storage - visuo-spatial processing, visuo-spatial storage - visuo-spatial processing and visuo-spatial storage - verbal processing).

An experimenter was present during the entire practice session to instruct and answer questions. After the practice sessions, the participant was allowed to complete the experimental trials independently, but the experimenter remained nearby in case of questions or problems. The experimenter could observe the participant's progress from outside the booth via a monitor that mirrored the participant's monitor and could see the participant through a window.

In each of the three sessions, participants performed each of the four combinations of storage-and-processing tasks in separate blocks, with the order of blocks randomly determined for each participant. Participants completed four blocks plus the task training in the first session, and four additional blocks in each of the remaining two sessions. Breaks were provided during each session in between blocks to mitigate potential fatigue effects. Each subsequent session after the first began with a brief description of the tasks to ensure that participants remembered what to do. At the end of the final session, participants were debriefed.

Each complex span task trial had a common construction: after presentation of a centrally displayed fixation cross for 1000 ms, a memory item was presented for 380-590 ms followed by a 500-ms pause and then a processing judgement. Participants observed one to-be-processed stimulus pair within a 4000-ms time window, corresponding approximately to the ratio between the time needed to perform a judgment and the time given to perform a judgment in Vergauwe et al.'s (2010) low cognitive load condition. Each letter pair appeared on screen until the participant's response, for up to 2667 ms, followed by a blank interval lasting the remainder of the 4000-ms period. Participants indicated their response ("yes" or "no") at any point during the 4000-ms period by clicking on buttons displayed centrally at the bottom of the screen and labeled accordingly. At the end of each complex span trial, serial reconstruction was prompted. For the tasks involving verbal memory, the 10 possible non-words appeared on screen in two rows of 5 non-words, ordered randomly to discourage spatially-mediated recall strategies. Participants were required to use the mouse and click on the memorized non-words in order of presentation. Once participants selected and clicked on a non-word, it turned grey and participants could not modify their response. All non-words disappeared simultaneously 500 ms after the participant had selected as many non-words at test as there were non-words presented at study. For the tasks involving visuo-spatial memory, the 10 possible locations appeared on screen and participants were required to use the mouse and click on the memory locations in order of presentation. Once participants selected and clicked on a location, it turned blue and participants could not modify their response. All locations disappeared 500 ms after the participant had selected as many locations at test as were presented at study. In all task conditions, participants were instructed to guess if a memory item was forgotten at test. After a 1000-ms blank interval, the next trial started.

Results and Discussion

Preliminary Manipulation Checks

To test our hypotheses, it is important that our tasks differed only in that verbal or visuo-spatial memory or processing is required. In designing our materials, we have ensured equivalencies in stimulus presentation, timings, and response options and methods, but it is also important to check that the processing tasks did not systematically differ in their difficulty and the time needed to respond. For analyzing response times, we excluded incorrect responses and trimmed responses more than 2.5 standard deviations from the mean, per participant (Grange, 2015). This resulted in the exclusion of 6% of otherwise eligible responses. Based on Jarrold et al. (2011), we expected the verbal and visuo-spatial processing tasks to produce similarly high accuracy rates and comparable response times. This was the case. Table 1 reports descriptive statistics for both processing tasks by session. Accuracies were consistently high across tasks and sessions. Response times decreased with each session for both tasks. On average, participants responded to rhyme judgments in 1307 ms (*SD* = 368) and to symmetry judgments in 1290 ms (*SD* = 349). Thus, as expected, verbal and visuo-spatial processing tasks produced similarly high accuracy rates and comparable response times. These values are comparable to those previously reported (Jarrold et al., 2011) and correspond to cognitive loads of 0.33 for rhyme judgments and 0.32 for symmetry judgments, following the proposed formula by (Barrouillet et al., 2007), that is, the ratio between the sum of the mean response times per processing phase, divided by the duration of the processing phase (i.e., inter-memoranda interval).

	Mean accuracy (SD)	Mean RT in ms (SD)
Experiment 1		
Rhyme judgments		
Session 1	0.96 (0.20)	1399 (362)
Session 2	0.95 (0.22)	1298 (365)
Session 3	0.95 (0.22)	1252 (361)
Symmetry judgments		
Session 1	0.97 (0.16)	1371 (346)
Session 2	0.98 (0.14)	1308 (340)
Session 3	0.97 (0.16)	1219 (345)
Experiment 2		
Rhyme judgments		
Session 1	0.97 (0.18)	1569 (376)
Session 2	0.97 (0.16)	1457 (354)
Session 3	0.98 (0.15)	1402 (348)
Symmetry judgments		
Session 1	0.98 (0.13)	1608 (349)
Session 2	0.99 (0.11)	1491 (332)
Session 3	0.99 (0.11)	1383 (351)

Note. *N*=19 in Experiment 1 (with one participant missing session 3). *N*=20 in Experiment 2 (with two participants missing sessions 2 and 3). Raw response times (RTs) were first trimmed per participant, removing responses more than 2.5 standard deviations from the mean.

Table 1. Descriptive statistics for processing tasks, by session, for Experiments 1 and 2.

Additionally, making use of the practice trials, we confirmed that processing and memory tasks had a disruptive effect on each other by checking that processing response times were increased and memory recall rates were reduced in complex span situations compared with single-task baselines taken in the first session. As expected, based on previous studies, memory recall decreased (baseline: M = 99% correct, complex span: M = 89%, both with 2-item lists) and processing response times slowed (for rhyme judgments, baseline M= 1358 ms and complex span M=1507 ms; for symmetry, baseline M= 1283 ms and complex span M= 1536 ms) when performed concurrently.

Finally, our adaptive procedure ensured that the complex span tasks were not too easy for each participant. List lengths increased by one item for participants who correctly recalled on average at least 80% of the memoranda in the preceding session, per complex span task. The number of participants who incremented list lengths in sessions 2 and 3 is listed in Supplementary materials 2. Most participants achieved at least 80% correct in Session 1 and list lengths increased accordingly (i.e., 15 out of 19 participants when verbal memory was combined with rhyme judgments, and 17 out of 19 participants for all other task combinations). In subsequent analyses, the three list lengths were coded as short, medium, and long, regardless of their absolute value, to be meaningful between participants and sessions.

Main Hypothesis Test

Figures 3A and 3B show plots of recall performance in each task, scored partially (proportion of items recalled correctly per list, referred to as *lenient scoring*; Figure 3A) or list-wise (proportion of whole lists recalled correctly, referred to as *strict scoring*; Figure 3B), both scores taking order information into account. Bayesian ANOVAs were run in RStudio (RStudio Team, 2020), using the Bayes factor package (Morey & Rouder, 2018) with default priors, following the method proposed by Rouder et al., 2012, with Processing task (rhyme vs. symmetry), Domain of storage (verbal vs. visuo-spatial) and Set size (short, medium, long) as within-subject variables. The code is publicly available on OSF (https://osf.io/mda6y/) and additional information can be found in Supplementary Materials 3. Patterns of results were comparable regardless of which scoring method we applied. Longer list lengths generally produced poorer recall (evidence for the main effect of Set size: $BF_{10} = 7.03x10^{43}$ and $BF_{10} = 9.27x10^{54}$ for lenient and strict scoring, respectively). With these stimuli, spatial locations were on average recalled more accurately than nonwords (evidence for the main effect of Domain of storage: $BF_{10} = 1.87x10^{29}$ and $BF_{10} = 1.06x10^{32}$ for lenient and strict scoring, respectively). The spread of accuracies observed with varying list lengths was wider for nonwords than spatial locations (evidence for the interaction between Domain of storage and Set size: $BF_{10} = 4.60x10^{11}$ and $BF_{10} = 8.93x10^{10}$ for lenient and strict scoring, respectively). Recall was slightly better on both tasks when paired with the symmetry judgment task (evidence for the main effect of Processing task: *BF*s ranged from an inconclusive 1 with strict scoring to a moderate 5 with lenient scoring).

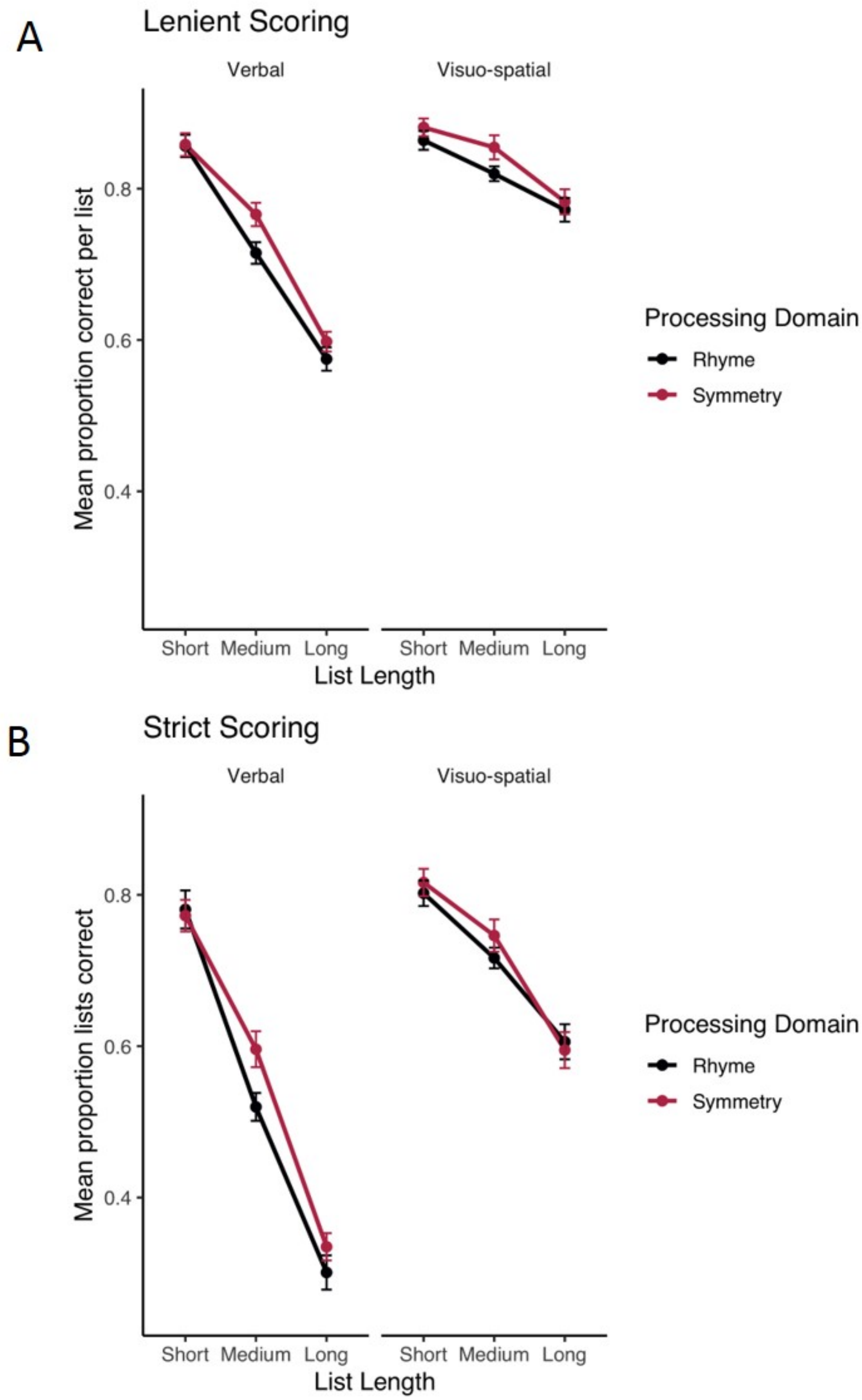

Figure 3. Mean proportional list recall (lenient scoring, Panel A) and Mean proportion of lists correctly recalled (strict scoring, Panel B) for each combination of storage domain (panels) and processing domain (colors) in Experiment 1. *N* = 19. Error bars are standard errors of the mean with the Cousineau-Morey (R. D. Morey, 2008) correction applied.

Importantly, to test whether a single, domain-general factor can account for working memory performance, or whether we need to propose a model that includes multiple domain-specific factors, we applied Bayesian state-trace analysis. Figure 4 shows state-trace plots separately for each participant, using lenient (Figure 4A) or strict scoring (Figure 4B). Visual inspection revealed a monotonic pattern, indicating that the domain of processing affects verbal and visuo-spatial memory performance similarly. To test whether a monotonic or a multi-factor solution better fit these data, we applied the $\hat{M}$ model, using strictly-scored responses (Cox & Kalish, 2019)[4]. $\hat{M}$ is a sample-wide statistic ranging from 0 - 1, and indicates the evidence favoring a unidimensional explanation for covariance between the verbal and visuo-spatial recall scores, as opposed to a multidimensional explanation. When $\hat{M}$ deviates above 0.5, it indicates evidence in favor of a single factor solution. The $\hat{M}$ for these data was 0.93. When the bulk of the posterior distribution of $\widehat{M}$ lies above 0.5, it is considered as strong evidence for a unidimensional explanation (see Supplementary Materials 4 for analyses of the data per session, showing the same results). The 95% highest density interval of the posterior distribution ranged from 0.75-0.99. This is very strong evidence in favor of a single-factor account of our data. Thus, the results of Experiment 1 clearly favored a single-factor account, both graphically and statistically.

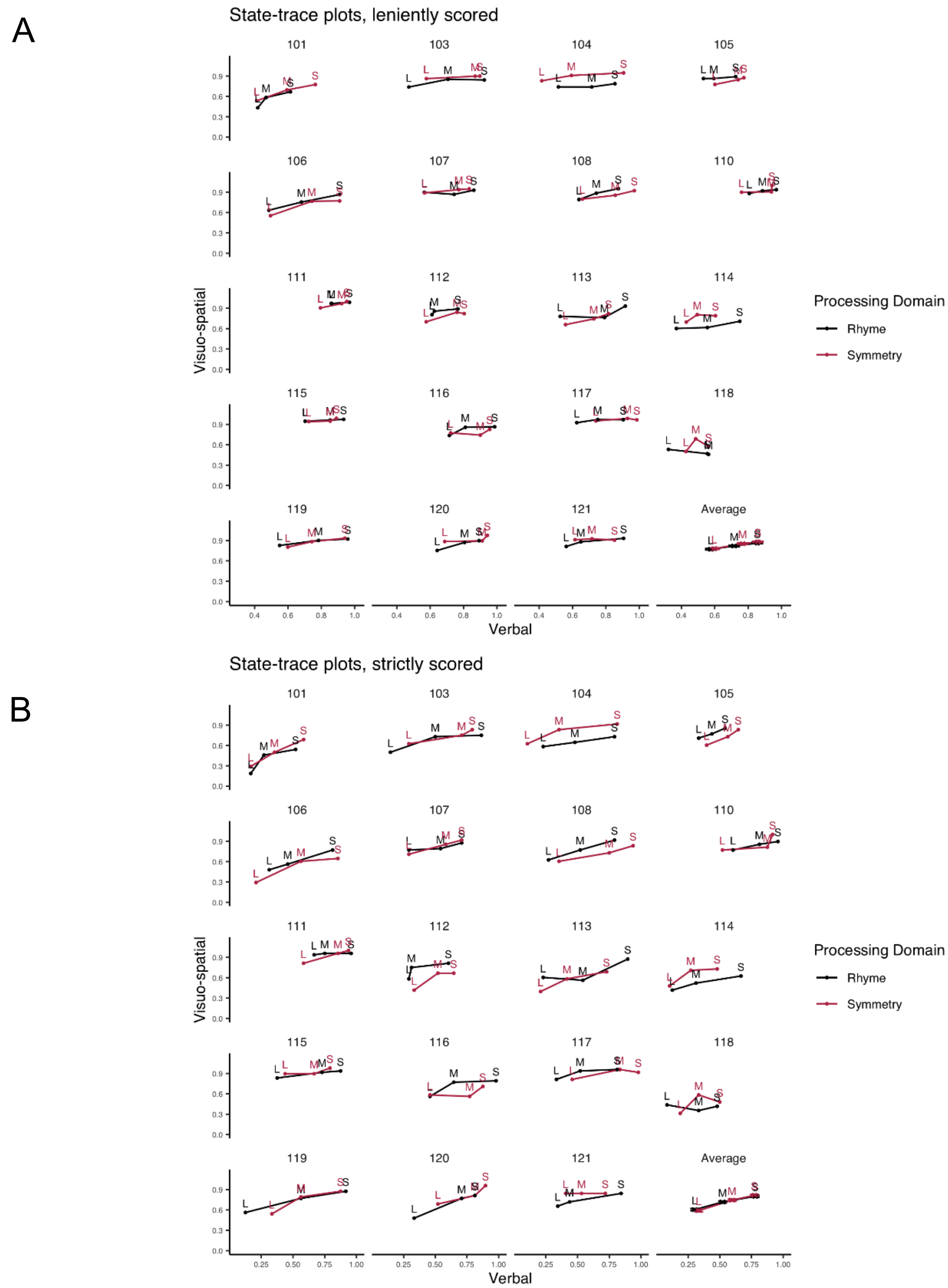

Figure 4. State-trace plots, with mean recall accuracy of verbal lists plotted against mean accuracy of visuo-spatial lists (Lenient scoring in Panel A, strict scoring in Panel B) in Experiment 1. Each panel represents a single participant, with averages in the final panel. *L*, *M*, and *S* stand for long, medium, and short lists, respectively, and refer to the lengths of the lists given to a participant within a session. Error bars on the *Average* panel are standard errors of the mean with the Cousineau-Morey (Morey, 2008) correction applied.

Given that our tasks were designed to limit participants to relying on phonological representation during nonword memory and on visuo-spatial representation during visuo-spatial memory, it is reasonable to conclude that only a single latent factor is necessary to support representation of phonological and visuo-spatial memoranda. However, one limitation of these results is the restricted range of list length effects, particularly when remembering visuo-spatial memoranda. We therefore repeated the experiment on a new sample of participants, for whom the lists started with 2, 4, or 6 items to increase the spread of recall scores in the three list lengths[3].

Experiment 2

Method

Participants

We recruited 24 participants. One participant dropped out of the experiment during the first session, and three participants were excluded after the first session because of below-threshold performance on at least one of the processing tasks. This resulted in N = 20 (3 males, 17 females), 20 – 54 years old (M = 29.35, SD = 9.54). Two participants were asked to come to the lab four instead of three times because of poor internet connection resulting in missing data during one of their sessions; they only completed the component that was missing in the fourth session. Two of the included participants completed only one session, because they could not negotiate a time for the remaining two with the researcher. Payment and inclusion criteria were kept the same as in Experiment 1. Sessions were again scheduled at least 24 hours apart, up to as much as 15 days apart. The study protocol was approved by the Ethics Committee of Cardiff University.

Materials and Procedure

The materials were the same as in Experiment 1, except that each participant started with list lengths of 2, 4, and 6 items in session 1. If accuracy was below 80% the list length remained the same. Otherwise, it increased to 3, 5, and 7 items in session 2, and similarly if performance was above 80%, the list length in session 3 rose to 4, 6, and 8. Due to the difference in list lengths, each session lasted for somewhat longer compared to Experiment 1. Each participant completed the 3 sessions in approximately 6.5h.

Results and Discussion

Preliminary Manipulation Checks

Like in Experiment 1, verbal and visuo-spatial processing tasks produced similarly high accuracy rates and comparable response times in Experiment 2. Following the same procedure to deal with processing task responses as described in Experiment 1, trimming resulted in the exclusion of ~4% of otherwise eligible responses. Mean accuracy rates and response times per session can be found in Table 1. Again, performance on the tasks was quite comparable. On average, participants responded to rhyme judgments in 1471 ms (SD = 365) and to symmetry judgments in 1487 ms (SD = 356). These values correspond to average cognitive loads of 0.37 for both kind of judgment, following the same formula as described in Experiment 1.

Furthermore, like in Experiment 1, processing and memory tasks had a disruptive effect on each other; memory recall decreased (baseline: M = 98% correct, complex span: M = 89%, both with 2-item lists) and processing response times slowed (for rhyme judgments, baseline M= 1549 ms and complex span M=1606 ms; for symmetry, baseline M= 1619 ms and complex span M= 1752 ms) when performed concurrently. Finally, our adaptive procedure ensured again that the complex span tasks were not too easy for each participant, and the number of participants who incremented list lengths in sessions 2 and 3 is listed in Supplementary materials 2.

Main Hypothesis Test

Figure 5 shows plots of recall performance in each task, scored leniently (Figure 5A) or strictly (Figure 5B). The same Bayesian ANOVAs were run as described in Experiment 1. The code is publicly available on our OSF page (https://osf.io/49cxt/) and additional information can be found in Supplementary Materials 3. Patterns of results were comparable regardless of which scoring method we applied. Longer list lengths generally produced poorer recall (evidence for the main effect of Set size: BF_{10} = 7.31x10^{92} and BF_{10} = 7.58x10^{102} for lenient and strict scoring, respectively). With these stimuli, spatial locations were recalled more accurately on average than nonwords (evidence for the main effect of Domain of storage: BF_{10} = 1.90x10^{5} and BF_{10} = 2.28x10^{8} for lenient and strict scoring, respectively). The spread of accuracies observed with varying list lengths was wider for nonwords than spatial locations (evidence for the interaction between Domain of storage and Set size: BF_{10} = 1693 and BF_{10} = 7835 for lenient and strict scoring, respectively). Excluding the effect of processing domain was favored in both analyses (evidence again the main effect of Processing task: *BF*s ranged from 5 with strict scoring to 7 with lenient scoring).

Figure 6 shows state-trace plots separately for each participant, using lenient (Figure 6A) or strict scoring (Figure 6B). Visual inspection shows indeed more spread in the visuo-spatial recall scores in Experiment 2, relative to those observed in Experiment 1. Moreover, visual inspection of the state-trace plots revealed again a monotonic pattern, indicating that the domain of processing affects verbal and visuo-spatial memory performance similarly. The $\widehat{M}$ (calculated using strict scores) for these data was

0.94 (see Supplementary Materials 4 for analyses of the data per session, showing the same results); the 95% highest density interval of the posterior distribution ranged from 0.83-0.99. Thus, as in Experiment 1, these values suggest that the data can be parsimoniously explained by assuming that a single factor supports storage of both verbal and visuo-spatial memoranda. Hence, across two carefully designed experiments, our results unambiguously favor a single, domain-general account of working memory performance.

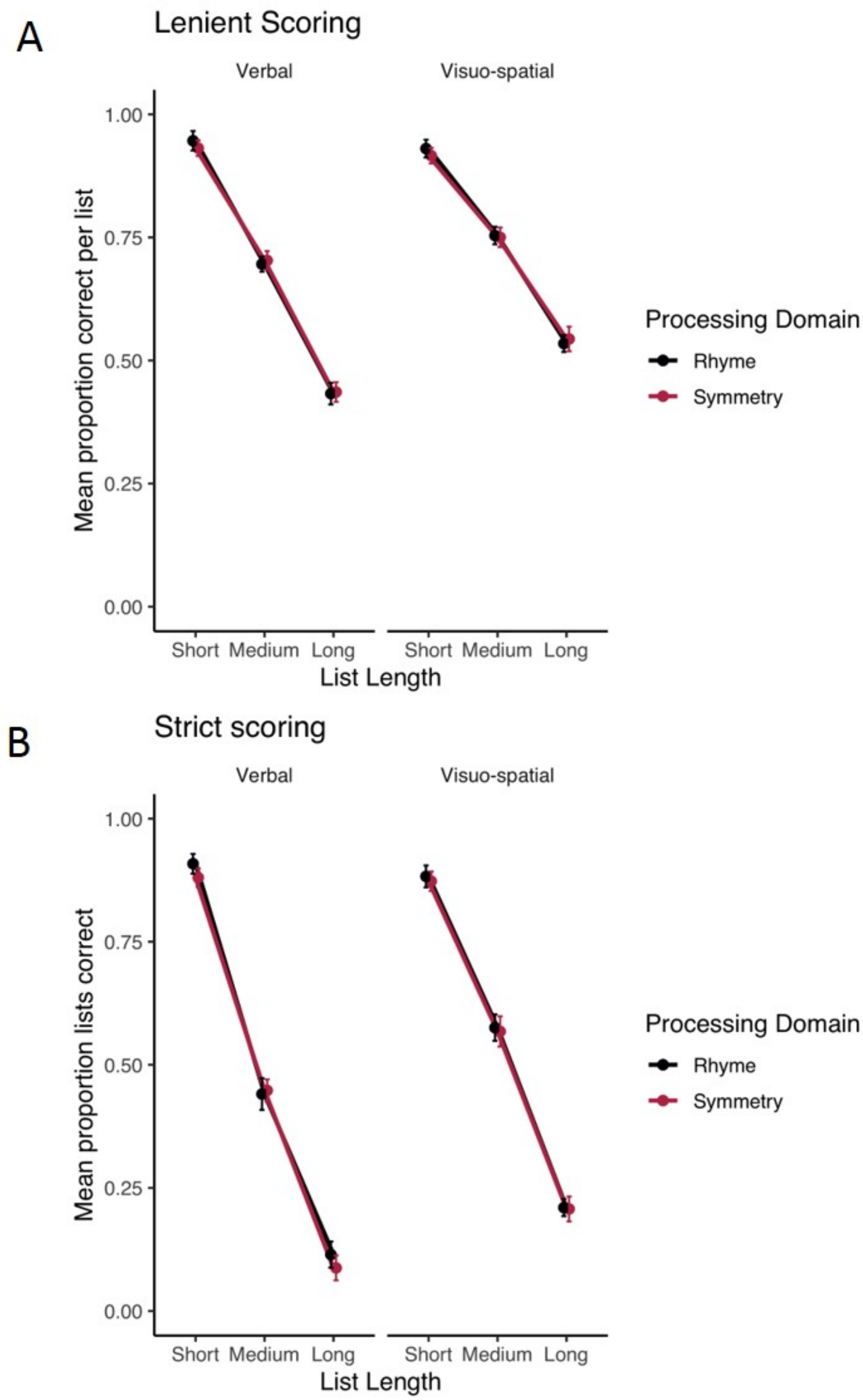

Figure 5. Mean proportional list recall (lenient scoring, Panel A) and Mean proportion of lists correctly recalled (strict scoring, Panel B) for each combination of storage domain (panels) and processing domain (colors) in Experiment 2. *N* = 20. Error bars are standard errors of the mean with the Cousineau-Morey (Morey, 2008) correction applied.

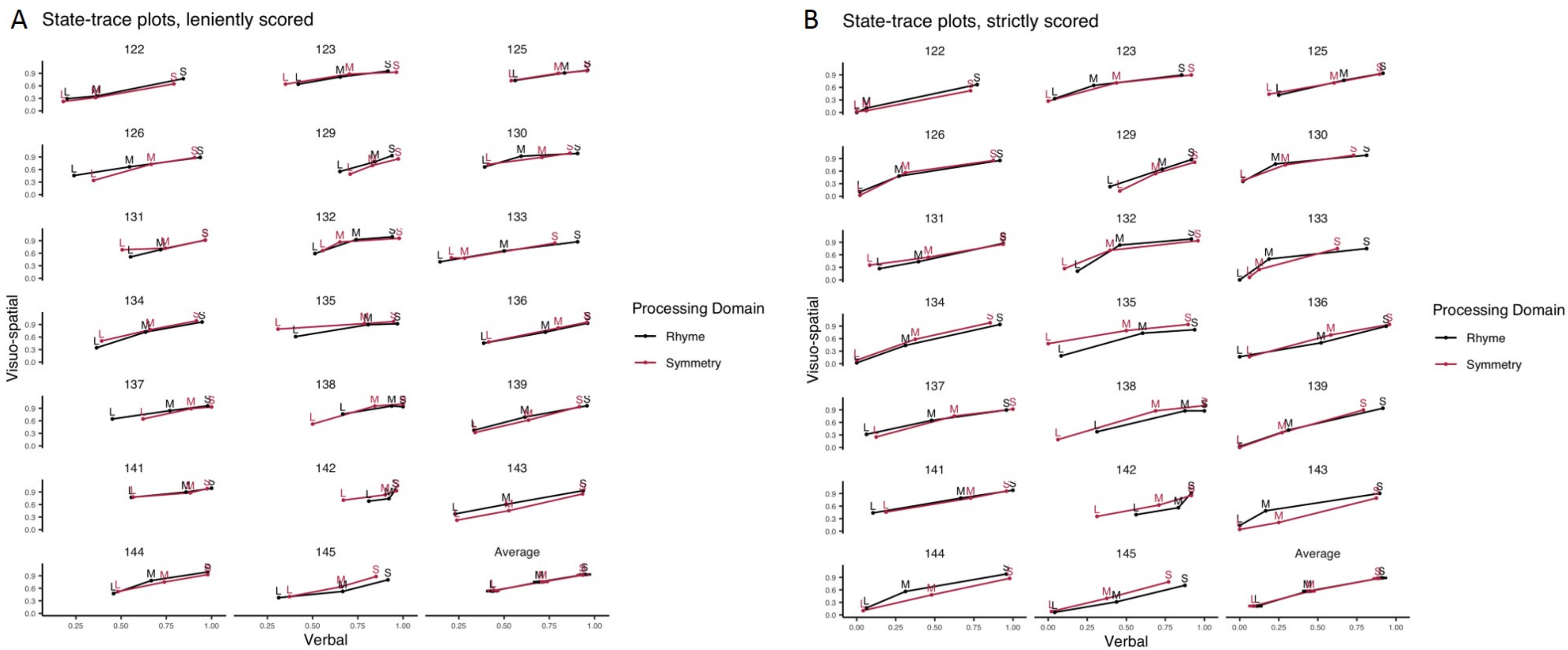

Figure 6. State-trace plots, with mean recall accuracy of verbal lists plotted against mean accuracy of visuo-spatial lists (lenient scoring in Panel A, strict scoring in Panel B) in Experiment 2. Each panel represents a single participant, with averages in the final panel. *L*, *M*, and *S* stand for long, medium, and short lists, respectively, and refer to the lengths of the lists given to a participant within a session. Error bars on the *Average* panel are standard errors of the mean with the Cousineau-Morey (Morey, 2008) correction applied.

General Discussion

Our study proposed a novel approach to a long-standing debate in psychology: the nature of the memory resources supporting multi-tasking performance. Using working memory theory and paradigms, we designed task situations that combine memory storage and processing requirements, and carefully manipulated the domain of storage and the domain of processing. We leveraged Bayesian state-trace analysis to test whether a single, domain-general factor can account for maintenance of verbal and visuo-spatial information, or whether we need to propose a working memory model that includes multiple domain-specific memory stores. The results of Experiment 1 clearly favor a single-factor account of working memory performance. We ran Experiment 2 to exclude alternative interpretations of the observed single-factor solution in Experiment 1 possibly arising from the limited spread of memory performance. Doing so, we established again that a single, domain-general factor can account for working memory performance and, thus, more generally, of multi-tasking performance.

Critically, our experiments were designed to highlight differences that would occur if there were separate short-term memory stores for verbal and visuo-spatial materials. Our memory tasks were designed (1) to limit participants to relying on phonological representation during nonword memory and on visuo-spatial representation during visuo-spatial memory, and (2) to minimize any differences beyond the difference of the domain of the memory representation by matching both the presentation and the recall phase of the tasks. Similarly, our processing tasks were designed (1) to tap into phonological processing during rhyme judgement and into visuo-spatial processing during symmetry judgement, and (2) to minimize any differences beyond the difference of the domain of the memory representation by matching processing materials, processing times, and processing accuracy as closely as possible. Doing so enabled us to isolate specifically the domain of the cognitive presentations involved in the different task versions, which was necessary to provide a stringently conservative test of domain-specific vs. domain-general accounts of working memory. One might argue that participants relied on episodic long-term memory to maintain the memory lists, rather than on working memory. However, it should be noted that our serial reconstruction memory task required the maintenance of both item and order information, and that the use of a closed set of ten non-words and ten spatial locations makes it highly unlikely that participants could simply rely on long-term memory and the arising familiarity to accomplish our task.

With our design, many outcomes were possible; one of them would support a strict domain-general account and all other outcomes would have supported domain-specific accounts. Had a multi-factor solution been favored by the data, our approach would not have allowed for uncovering the exact nature of the multi-dimensionality. The most straightforward interpretation would have been in terms of verbal vs. visuo-spatial short-term buffers, but in principle, a diverging pattern could also have pointed to other qualitative differences between verbal and visuo-spatial storage, such as the use of verbal rehearsal for one material but not for the other, or storage of multiple items as a set of distinct items vs. integrated configurations. Equating verbal and visuo-spatial tasks as much as possible, except for the domain of the representations, our results unequivocally show that only a single latent factor is necessary to support the representation of phonological and visuo-spatial memoranda in working memory, thereby ruling out an entire class of domain-specific accounts and providing a conclusive answer to this persistent question.

Because we designed the experiments so that the tasks are equated as much as possible, one could argue that our findings are specific to the tasks that we have used and therefore may not generally rule out multi-dimensional views. We agree that by focusing on the domain of the representations in our study, we provided one of the most conservative tests of domain-specific vs. domain-general accounts of short-term storage to date. We consider this a strength of this study. Notably, our design choices regarding memoranda, processing materials and tasks, and complex span procedure all have precedent and have been used before by proponents of both domain-specific and domain-general views on working memory (e.g., Bayliss et al., 2003; Jarrold et al., 2011; Morey & Mall, 2012; Morey & Miron, 2016; Shah & Miyake, 1996; Vergauwe et al., 2010, 2012). However, previous studies may not have entirely isolated representation of verbal and visuo-spatial information. Though each of the memory and processing stimuli we chose have been used in previous work, to our knowledge they have not previously been inserted into procedures in which timing, numbers of response options, and response mode (Thalmann & Oberauer, 2017), in either the processing task or during recall, were as closely matched as in our paradigm[6]. Therefore, even though our strong conclusion is based on only one type of verbal vs. visuo-spatial memory material and only one type of verbal vs. visuo-spatial processing tasks, it seems highly unlikely that the pattern would be different in future studies that carefully manipulate domain of representation while using other memory materials or processing tasks. It would be exciting to see a program of research that systematically isolated the many factors that usually vary between verbal and visuo-spatial complex span tasks, which could potentially reveal the most likely sources for domain-specific effects that have been observed in several previous studies. In a similar vein, our conclusion is based on only one type of task, i.e., the complex span. Even though the complex span task is one of the tasks that has been used most frequently to examine the nature of working memory resources (Bayliss et al., 2003; Jarrold et al., 2011; Kane et al., 2004; Shah & Miyake, 1996; Thalmann & Oberauer, 2017; Vergauwe et al., 2010, 2012), future research should examine to what extent the current pattern extends to other working memory tasks, with special attention to tasks that have previously shown domain-specific effects. Our evidence points away from domain-specific short-term stores as a likely source for these effects.

In contrast to previous studies finding patterns of (partly) selective interference, we observed that verbal and visuo-spatial memory performance were both slightly worse when combined with rhyme judgment than when combined with symmetry judgment in Experiment 1, and they were not impacted by the domain of concurrent processing in Experiment 2 (see Figures 3 and 5). On the one hand, observing so little variability of the effects of processing domain on recall of verbal and visuo-spatial memoranda may have limited our ability to detect a multidimensional outcome in the state-trace analysis. From a theoretical perspective, the contrasting processing stimuli we chose were intended to maximize differential effects, if working memory draws on separate verbal and visuo-spatial resources. One way, then, to account for the lack of selective interference would be to assume that the rhyme and symmetry tasks are both equally verbal and visuo-spatial in nature. This is, however, highly implausible. Even though some visual processing may be involved in the rhyme judgment task and some verbal processing may be involved in the symmetry judgment task, the two tasks differ clearly in the representations needed to perform the task; rhyme judgment requires the comparison of phonological representations whereas symmetry judgment requires the comparison of visuo-spatial representations. For our approach to work, it is not necessary that the two processing tasks are entirely distinct from each other. However, our approach does depend on accepting that the rhyme and symmetry processing tasks rely predominantly on phonological vs. visuo-spatial representations, respectively. Based on common assumptions in the field of cognitive psychology, it appears reasonable to assume that judging whether two stimuli rhyme requires the comparison of phonological, speech-related representations, whereas judging whether two stimuli share the same axis of symmetry requires comparison of visuo-spatial, space-related representations. It is important to stress the strength of our state-trace approach here; whereas the observed pattern of interference could have been “reasoned away” by proposing post hoc, multiple-resource accounts, that is simply impossible for the state-trace analysis results which unequivocally show that a single factor is sufficient to account for the observed memory behavior.

Importantly, the fact that processing domain impacted recall performance only slightly should not be taken as reflecting a total absence of resource-sharing between the concurrent processing and storage tasks. Indeed, we also demonstrated that processing response times increased and memory recall rates decreased in complex span situations compared with single-task baselines, consistently with previous research (Bayliss et al., 2003; Jarrold et al., 2011)[5]. The clear absence of the expected interaction between processing and storage domains showing selective interference, combined with the equality of accuracy and speed on the two processing tasks denoting their nearly equivalent cognitive load, further supports the conclusion that short-term storage of verbal and visuo-spatial memoranda is supported by a single latent factor.

Although our findings are clearly inconsistent with proposing separate verbal and visuo-spatial short-term memory stores, working memory is not necessarily a unidimensional cognitive system overall. It is indeed important to acknowledge that our findings rule out a narrow assumption of the domain-specific view: the domain-specificity of temporary storage for verbal and spatial representations. However, this specific tenet is theoretically highly influential, and continues to guide many researchers interested in working memory and multi-tasking across several areas of psychology (cognitive, clinical, educational, and developmental). Therefore, our test was a crucial one.

What are the implications for models of working memory? After all, some versions of the multiple-component model have proposed the involvement of domain-general resources in addition to domain-specific storage resources, and some versions of more unitary views have also proposed the existence of domain-specific storage resources (at least for verbal representations). Concerning the multiple-component model, it is important to note that although different versions of model diverge on some crucial assumptions related to the exact nature of working memory resources underlying processing and storage activities, they all converge on the existence of domain-specific temporary storage for verbal and spatial representations. Different versions of the multiple-component diverge on the extent to which processing and storage can interfere with each other because (1) some versions propose the existence of domain-general resources supporting processing and storage in addition to domain-specific storage resources (e.g., Baddeley, 2000; Baddeley & Hitch, 1974) whereas others explicitly reject this notion (e.g., Baddeley & Logie, 1999; Cocchini et al., 2002; Duff & Logie, 2001), and (2) some versions propose that the domain-specific storage components are also involved in processing activities (e.g., Baddeley & Logie, 1999) whereas others explicitly reject this notion (e.g., Cocchini et al., 2002; Duff & Logie, 2001). However, what all these versions of the multiple-component model have in common is that they all propose more than one resource for working memory storage of verbal and visuo-spatial representations. Therefore, regardless of the nature of their assumptions other than the domain-specificity of temporal storage of verbal and visuo-spatial representations, all versions of the multiple-component model would have resulted in diverging memory performance traces, which is not what we observed. In the same way, versions of domain-general views that include one or more domain-specific stores for verbal and/or visuo-spatial representations in addition to domain-general storage are contradicted by our findings of a monotonic relationship. That is, any model of working memory that views working memory as a system, or a set of processes, holding mental representations temporarily available for use in ongoing thought and action (i.e., the vast majority of existing models), and proposes more than one resource supporting temporary storage of verbal and visuo-spatial representations is contradicted by our findings. Our findings do not deny the existence of domain-specific effects in working memory, but they do rule out explanations of those effects in terms of domain-specific stores for verbal and visuo-spatial representations.

One possible caveat in regard to our conclusion arises from neuropsychological and neuroimaging studies suggesting that verbal and visuo-spatial memoranda might be stored in different brain regions (e.g., Emrich et al., 2013; Harrison & Tong, 2009; Serences et al., 2009). While domain-specific effects are often found in these studies, we think that these findings do not necessarily point to separate buffers in the brain. Indeed, the described dissociations in patient cases have recently been called into question and alternative accounts have been proposed that do not rely on domain-specific storage buffers (Buchsbaum & D'Esposito, 2008, 2019; Caplan et al., 2012;

Caplan & Waters, 1999; Morey, 2018; Morey et al., 2019, 2020). Furthermore, although memoranda can be decoded from sensory regions, these studies rarely are concerned with multi-tasking situations and, more importantly, several studies have shown that memoranda can also be decoded from other, non-sensory regions (e.g., Bettencourt & Xu, 2016; Christophel et al., 2012; Ester et al., 2015). Overall, in line with what was argued by Morey (2018; Morey et al., 2019), most of these findings could be accounted for by a domain-general working memory system in combination with a sensory-motor integration account, without the need to assume separate, domain-specific short-term stores.

Debate about working memory components may naturally shift toward testing whether perceptual- or response-led factors may drive the consistent domain-specific effects that certainly appear in several studies in the working memory literature (see Oberauer et al., 2018, for an overview). Similarly, a recent study showed that there is much more evidence for a separation of resources in terms of response modalities (manual vs. vocal) than in terms of verbal vs. visuo-spatial processing codes when it comes to multi-tasking performance (Bruning et al., 2020). Models of working memory and multi-tasking may then be updated to better account for these effects.

Conclusion

To conclude, across two experiments, we fully crossed processing and storage tasks that were designed to contrast only the domain of representation (verbal vs. visuo-spatial) in order to test whether multi-tasking is supported by multiple, specialized short-term storage resources, as commonly presumed. The results that we obtained in this strict, theory-based testing environment unambiguously favor a simple, single-factor account of working memory performance, leading to only one conclusion: multi-tasking in working memory is supported by a single, domain-general cognitive resource for briefly representing information.

Footnotes

Footnote 1. This view on working memory corresponds to what Cowan (2017) referred to as the "storage-and-processing definition" of working memory, proposing working memory as a cognitive workspace for both short-term storage and processing, as opposed to short-term memory only having a storage function.

Footnote 2. With the current approach, evidence for a unidimensional solution would come from finding a high $\widehat{M}$, i.e., a high level of evidence in favor of a single factor solution, as opposed to not being able to reject the null hypothesis when it comes to the interaction between processing domain and storage domain in the traditional selective interference studies (i.e., p>.05).

Footnote 3. We deviated from our preregistration in using this method; our pre-registration specified that we would use the Prince et al. (2012) method. However, in between planning this study and carrying out the analysis, the Cox and Kalish paper presented a new method for computing a sample-wide statistic for conveniently quantifying evidence for monotonicity. Prince et al.'s method assessed monotonicity per individual in the sample, yielding a group summary Bayes factor which was interpretable only if participants' patterns were homogenous (Davis-Stober et al., 2016). Prince et al.'s method was also designed for analysis of bimodal responses. Serial reconstruction accuracy can be validly computed in multiple ways (e.g., whole-list correct or not, proportion of list correct, etc.) so being able to assess evidence for monotonicity for multiple methods of scoring allowed us to check the robustness of our findings. Note that we did not carry out an analysis using Prince et al.'s method (their R package was no longer up-to-date at the time of our analysis). Originally, we had planned to start with an initial sample size of 20 participants, and to implement sequential hypothesis testing after acquiring the data of 20 participants. Our plan was to verify whether we had sample-level evidence in the form of a Bayes factor of 10 (coming from the group summary output from Prince et al.'s *R* package) or more favoring either the multi-factor or single-factor solution. If we had conclusive data, data collection would stop. If we had not reached a Bayes factor of 10, we had planned to continue collecting data in batches of 4-6 new participants, re-analyzing the data in between each new batch until we reached a Bayes factor of 10 favoring either hypothesis or reach a total N=45. Switching to the Cox and Kalish method entailed the use of the outcome statistic $\widehat{M}$ which assesses the degree to which a dataset is consistent with a monotonic state trace at the group level. Because the analyzed sample of 19 participants in Experiment 1 provided decisive support for a unidimensional explanation, we did not further increase the sample size. Instead, we used our remaining resources to run Experiment 2, resulting again in decisive support for a unidimensional explanation.

Footnote 4. We also ran the $\widehat{M}$ model on leniently-scored values, with comparable results ($\widehat{M}$= 0.95; 95% HDI range = 0.82-0.99).

Footnote 5. The observation that memory performance suffered from adding a concurrent processing task demonstrates that the memory representations involved in our paradigm were vulnerable to interference. This is relevant for views on working memory that distinguish between a short-duration, fragile store and a longer-duration, more robust store (e.g., Sligte et al., 2008). Indeed, one could argue that the use of the complex span paradigm resulted in rather long trials, thereby tapping rather into the more durable store. However, the fact that we did observe interference from concurrent processing appears to speak against that idea.

Footnote 6. We made an explicit effort to isolate and manipulate only the domain of cognitive representations involved in the different tasks. Concerning the memory tasks, the domain of presentation was manipulated through the use of aurally-presented non-words vs. visually-presented spatial locations as memory items, while using the same presentation durations, the same sequential presentation mode, the same number of possible memory items, and the same serial reconstruction procedure with the same response mode. Using comparable verbal and visuo-spatial memory stimuli, Morey and Miron (2016) observed similar performance between verbal and visuo-spatial serial order reconstruction tasks for 5-item sequences. Concerning the processing tasks, the domain of presentation was manipulated through the requirement to make a rhyme vs. symmetry judgment, while using the same stimuli (two simultaneously presented letters), the same durations, the same number of response options, and the same response modes. In the visuo-spatial processing task, letters were inverted to discourage reading. Using comparable verbal and visuo-spatial processing stimuli, Jarrold et al. (2011) observed equivalent accuracy rates and response times between rhyme and symmetry judgments.

Acknowledgements

This work was completed with support from the British Academy (BA/Leverhulme small research grant SRG18R1\180622 to C.C.M.) and from the Swiss National Science Foundation (Grant PCEFP1_181141 to E.V.). Thanks are also due to Greg Cox for sharing scripts for computing $\widehat{M}$ and for support with running these analyses.

Author Contributions

The study was conceptualized by C.C.M. and E.V. The experiments were designed by C.C.M., E.V., and C.v.B. The experiment was programmed by C.v.B. The data were acquired by R.K. The data were analyzed by C.C.M. and R.K. The data were interpreted by C.C.M., R.K., E.V., and C.v.B. The manuscript was drafted by C.C.M. and E.V., with critical input from all authors.

References

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

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In *Recent advances in learning and motivation Volume VIII* (Vol. 8, pp. 47–89). Academic Press.

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2019). From short-term store to multicomponent working memory: The role of the modal model. *Memory & Cognition*, *47*(4), 575–588. https://doi.org/10.3758/s13421-018-0878-5

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge University Press.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon project. *Behavior Research Methods*, *39*(3), 445–459.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83–100. https://doi.org/10.1037/0096-3445.133.1.83

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 570–585. https://doi.org/10.1037/0278-7393.33.3.570

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review*, *118*(2), 175–192. https://doi.org/10.1037/a0022324

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*(1), 71–92. https://doi.org/10.1037/0096-3445.132.1.71

Berry, E. D. J., Allen, R. J., Waterman, A. H., & Logie, R. H. (2019). The effect of a verbal concurrent task on visual precision in working memory. *Experimental Psychology*, *66*(1), 77. https://doi.org/10.1027/1618-3169/a000428

Bettencourt, K. C., & Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience*, *19*(1), 150-+. https://doi.org/10.1038/nn.4174

Bruning, J., Muckstein, M., & Manzey, D. (2020). Multitasking strategies make the difference: Separating processing-code resources boosts multitasking efficiency when individuals prefer to interleave tasks in free concurrent dual tasking. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(12), 1411–1433. https://doi.org/10.1037/xhp0000865

Buchsbaum, B. R., & D'Esposito, M. (2008). The search for the phonological store: From loop to convolution. *Journal of Cognitive Neuroscience*, *20*(5), 762–778. https://doi.org/10.1162/jocn.2008.20501

Buchsbaum, B. R., & D'Esposito, M. (2019). A sensorimotor view of verbal working memory. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *112*, 134–148. https://doi.org/10.1016/j.cortex.2018.11.010

Camos, V., Mora, G., Offtinger, A.-L., Mariz Elsig, S., Schneider, P., & Vergauwe, E. (in press). Does semantic long-term memory impact refreshing in verbal working memory? *Journal of Experimental Psychology: Learning Memory and Cognition*.

Caplan, D., Waters, G., & Howard, D. (2012). Slave systems in verbal short-term memory. *Aphasiology*, *26*(3–4), 279–316. https://doi.org/10.1080/02687038.2011.642795

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*(1), 77–126. https://doi.org/10.1017/S0140525X99001788

Christophel, T. B., Hebart, M. N., & Haynes, J.-D. (2012). Decoding the Contents of Visual Short-Term Memory from Human Visual and Parietal Cortex. *Journal of Neuroscience*, *32*(38), 12983–12989. https://doi.org/10.1523/JNEUROSCI.0184-12.2012

Cocchini, G., Logie, R. H., Sala, S. D., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition*, *30*, 1086–1095.

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences*, *3*(3), 115–120.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786.

Cowan, N. (1995). *Attention and memory: An integrated framework.* Oxford University Press.

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

Cox, G. E., & Kalish, M. L. (2019). Dial Mˆ for monotonic: A kernel-based Bayesian approach to state-trace analysis. *Journal of Mathematical Psychology*, *90*, 100–117. https://doi.org/10.1016/j.jmp.2019.02.002

Davis-Stober, C. P., Morey, R. D., Gretton, M., & Heathcote, A. (2016). Bayes factors for state-trace analysis. *Journal of Mathematical Psychology*, *72*, 116–129.

Duff, S. C., & Logie, R. H. (2001). Processing and storage in working memory span. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *54A*(1), 31–48. https://doi.org/10.1080/02724980042000011

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, *95*(1), 91–101. https://doi.org/10.1037/0033-295X.95.1.91

Emrich, S. M., Riggall, A. C., LaRocque, J. J., & Postle, B. R. (2013). Distributed Patterns of Activity in Sensory Cortex Reflect the Precision of Multiple Items Maintained in Visual Short-Term

Memory. *Journal of Neuroscience*, *33*(15), 6516–6523. https://doi.org/10.1523/JNEUROSCI.5732-12.2013

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. https://doi.org/10.1037/0096-3445.128.3.309

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron*, *87*(4), 893–905. https://doi.org/10.1016/j.neuron.2015.07.013

Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press.

Foley, J. A., Kaschel, R., Logie, R. H., & Della Sala, S. (2011). Dual-task performance in Alzheimer's disease, mild cognitive impairment, and normal ageing. *Archives of Clinical Neuropsychology*, *26*(4), 340–348. https://doi.org/10.1093/arclin/acr032

Grange, J. (2015). *trimr: An Implementation of Common Response Time Trimming Methods* (1.0.1) [Computer software]. https://cran.r-project.org/web/packages/trimr/index.html

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. https://doi.org/10.1038/nature07832

Jarrold, C., Tam, H., Baddeley, A. D., & Harvey, C. E. (2010). The nature and position of processing determines why forgetting occurs in working memory tasks. *Psychonomic Bulletin & Review*, *17*(6), 772–777. https://doi.org/10.3758/PBR.17.6.772

Jarrold, C., Tam, H., Baddeley, A. D., & Harvey, C. E. (2011). How does processing affect storage in working memory tasks? Evidence for both domain-general and domain-specific effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 688–705. https://doi.org/10.1037/a0022527

Kahneman, D. (1973). *Attention and effort*. Prentice Hall.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Logie, R. H., Della Sala, S., Wynn, V., & Baddeley, A. D. (2000). Visual similarity effects in immediate verbal serial recall. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *53*(3), 626–646. https://doi.org/10.1080/027249800410463

Morey, C. C. (2018). The case against specialized visual-spatial short-term memory. *Psychological Bulletin*, *144*(8), 849–883. https://doi.org/10.1037/bul0000155

Morey, C. C., & Mall, J. T. (2012). Cross-domain interference costs during concurrent verbal and spatial serial memory tasks are asymmetric. *Quarterly Journal of Experimental Psychology*, *65*(9), 1777–1797. https://doi.org/10.1080/17470218.2012.668555

Morey, C. C., & Miron, M. D. (2016). Spatial sequences, but not verbal sequences, are vulnerable to general interference during retention in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(12), 1907–1918. https://doi.org/10.1037/xlm0000280

Morey, C. C., Rhodes, S., & Cowan, N. (2019). Sensory-motor integration and brain lesions: Progress toward explaining domain-specific phenomena within domain-general working memory. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *112*, 149–161. https://doi.org/10.1016/j.cortex.2018.11.030

Morey, C. C., Rhodes, S., & Cowan, N. (2020). Co-existing, contradictory working memory models are ready for progressive refinement: Reply to Logie. *Cortex*, *123*, 200–202. https://doi.org/10.1016/j.cortex.2019.11.012

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial for Quantitative Methods in Psychology*, *4*, 61–64.

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes factors for common designs* (0.9.12) [Computer software]. http://cran.r-project.org/web/packages/BayesFactor/index.html

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, *86*(3), 214–255. https://doi.org/10.1037/0033-295X.86.3.214

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*(1), 44–64. https://doi.org/10.1016/0010-0285(75)90004-3

Oberauer, K. (2013). The focus of attention in working memory-from metaphors to mechanisms. *Frontiers in Human Neuroscience*, *7*, 673. https://doi.org/10.3389/fnhum.2013.00673

Oberauer, K., Farrell, S., Jarrold, C., Pasiecznik, K., & Greaves, M. (2012). Interference between maintenance and processing in working memory: The effect of item–distractor similarity in complex span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 665–685. https://doi.org/10.1037/a0026337

Oberauer, K., & Lange, E. B. (2008). Interference in verbal working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration. *Journal of Memory and Language*, *58*(3), 730–745. https://doi.org/10.1016/j.jml.2007.09.006

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958. https://doi.org/10.1037/bul0000153

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19*(5), 779–819. https://doi.org/10.3758/s13423-012-0272-4

Prince, M., Brown, S., & Heathcote, A. (2012). The design and analysis of state-trace experiments. *Psychological Methods*, *17*(1), 78–99. https://doi.org/10.1037/a0025809

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*(3), 164–171. psyh. https://doi.org/10.1027/1015-5759/a000123

Repovs, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience*, *139*(1), 5–21. https://doi.org/10.1016/j.neuroscience.2005.12.061

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

RStudio Team. (2020). *RStudio: Integrated Development for R*. RStudio, Inc. http://www.rstudio.com/

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*(2), 207–214. https://doi.org/10.1111/j.1467-9280.2009.02276.x

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4–27. https://doi.org/10.1037/0096-3445.125.1.4

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PLoS ONE*, *3*(2), e1699. https://doi.org/10.1371/journal.pone.0001699

Stephens, R. G., Matzke, D., & Hayes, B. K. (2019). Disappearing dissociations in experimental psychology: Using state-trace analysis to test for multiple processes. *Journal of Mathematical Psychology*, *90*, 3–22. https://doi.org/10.1016/j.jmp.2018.11.003

Thalmann, M., & Oberauer, K. (2017). Domain-specific interference between storage and processing in complex span is driven by cognitive and motor operations. *Quarterly Journal of Experimental Psychology*, *70*(1), 109–126. https://doi.org/10.1080/17470218.2015.1125935

Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource? *Psychological Science*, *21*(3), 384–390. https://doi.org/10.1177/0956797610361340

Vergauwe, E., Dewaele, N., Langerock, N., & Barrouillet, P. (2012). Evidence for a central pool of general resources in working memory. *Journal of Cognitive Psychology*, *24*(3), 359–366. https://doi.org/10.1080/20445911.2011.640625 von Bastian, C. C., Locher, A., & Ruflin, M. (2013). Tatool: A Java-based open-source programming framework for psychological studies. *Behavior Research Methods*, *45*(1), 108–115. https://doi.org/10.3758/s13428-012-0224-y

von Bastian, C. C., & Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. *Journal of Memory and Language*, *69*(1), 36–58. https://doi.org/10.1016/j.jml.2013.02.002

Wickens, C. D. (1980). The structure of attentional resources. In *Attention and Performance VIII* (R. Nickerson, pp. 239–257). Lawrence Erlbaum.

Wickens, C. D. (1984). Processing resources in attention. In *Varieties of attention* (R. Parasuraman&D.R. Davies, pp. 63–102). Academic Press.

Tables

Table 1. Descriptive statistics for processing tasks, by session, for Experiments 1 and 2.

	Mean accuracy (SD)	Mean RT in ms (SD)
Experiment 1		
Rhyme judgments		
Session 1	0.96 (0.20)	1399 (362)
Session 2	0.95 (0.22)	1298 (365)
Session 3	0.95 (0.22)	1252 (361)
Symmetry judgments		
Session 1	0.97 (0.16)	1371 (346)
Session 2	0.98 (0.14)	1308 (340)
Session 3	0.97 (0.16)	1219 (345)
Experiment 2		
Rhyme judgments		
Session 1	0.97 (0.18)	1569 (376)
Session 2	0.97 (0.16)	1457 (354)
Session 3	0.98 (0.15)	1402 (348)
Symmetry judgments		
Session 1	0.98 (0.13)	1608 (349)
Session 2	0.99 (0.11)	1491 (332)
Session 3	0.99 (0.11)	1383 (351)

Note. N=19 in Experiment 1 (with one participant missing session 3). N=20 in Experiment 2 (with two participants missing sessions 2 and 3). Raw response times (RTs) were first trimmed per participant, removing responses more than 2.5 standard deviations from the mean.

Figure captions

Figure 1. Schematic illustration of multidimensional (upper panel) and unidimensional (lower panel) accounts of working memory. The upper panel shows a multidimensional account, proposing two separate, domain-specific stores (one for verbal information, and one for visuo-spatial information; implemented in our study by using verbal vs. visuo-spatial memory tasks: memory for nonwords vs. memory for locations, respectively), that are both affected by memory load (implemented in our study by using memory lists of varying length) and differently affected by the domain involved in a concurrent processing task (implemented in our study by using verbal vs. visuo-spatial processing tasks: rhyme vs. symmetry judgment). If multiple, domain-specific latent factors underlie working memory storage, diverging hypothetical memory performance traces as shown in (A) should be observed, when plotting verbal memory performance (memory accuracy for nonwords, in our study) and visuo-spatial memory performance (memory accuracy for locations, in our study) against each other, as a function of Processing domain (dark vs. magenta line) and Memory load (different points in the graph). The dashed lines in (A) represent the selective interference that would reflect no dual-task cost (i.e., constant performance) in cross-domain combinations across the increasing set sizes. We do not expect performance on either task to sit near these dashed lines; their presence helps to show that many possible patterns (including asymmetric ones) short of complete absence of interference could reflect the operation of multiple latent factors. The lower panel shows a unidimensional account, proposing a single, domain-general store for both verbal and visuo-spatial information), that is affected by memory load and domain involved in a concurrent processing task. If a single, latent factor supports working memory storage, a monotonic pattern as shown in

(B) should be observed when plotting verbal memory performance and visuo-spatial memory performance against each other, as a function of Processing domain (dark vs. magenta line) and Memory load (different points in graph).

Figure 2. Panel A: Two processing tasks requiring decisions based on visuo-spatial or phonological aspects of the stimuli. In the symmetry task, participants judged whether two inverted, mirrored letters shared the same axis of symmetry (dotted lines are shown to denote each letter's axis of symmetry; these were not shown during the task). In the rhyme task, participants judged whether the pronunciation of upright letters rhymed. Examples of true and false pairings are provided for both tasks. Panel B: Events in a trial procedure combining visuo-spatial memoranda with rhyme processing judgments. In this example, both rhyme stimuli should elicit "No" responses. Participants completed blocks during each session fully crossing verbal and visuo-spatial memoranda with verbal and visuo-spatial processing.

Figure 3. Mean proportional list recall (lenient scoring, Panel A) and Mean proportion of lists correctly recalled (strict scoring, Panel B) for each combination of storage domain (panels) and processing domain (colors) in Experiment 1. $N = 19$. Error bars are standard errors of the mean with the Cousineau-Morey(R. D. Morey, 2008) correction applied.

Figure 4. State-trace plots, with mean recall accuracy of verbal lists plotted against mean accuracy of visuo-spatial lists (Lenient scoring in Panel A, strict scoring in Panel B) in Experiment 1. Each panel represents a single participant, with averages in the final panel. *L*, *M*, and *S* stand for long, medium, and short lists, respectively, and refer to the lengths of the lists given to a participant within a session. Error bars on the *Average* panel are standard errors of the mean with the Cousineau-Morey (Morey, 2008) correction applied.

Figure 5. Mean proportional list recall (lenient scoring, Panel A) and Mean proportion of lists correctly recalled (strict scoring, Panel B) for each combination of storage domain (panels) and processing domain (colors) in Experiment 2. $N = 20$. Error bars are standard errors of the mean with the Cousineau-Morey (Morey, 2008) correction applied.

Figure 6. State-trace plots, with mean recall accuracy of verbal lists plotted against mean accuracy of visuo-spatial lists (Lenient scoring in Panel A, strict scoring in Panel B) in Experiment 2. Each panel represents a single participant, with averages in the final panel. *L*, *M*, and *S* stand for long, medium, and short lists, respectively, and refer to the lengths of the lists given to a participant within a session. Error bars on the *Average* panel are standard errors of the mean with the Cousineau-Morey (Morey, 2008) correction applied.

Figures

Figure 1

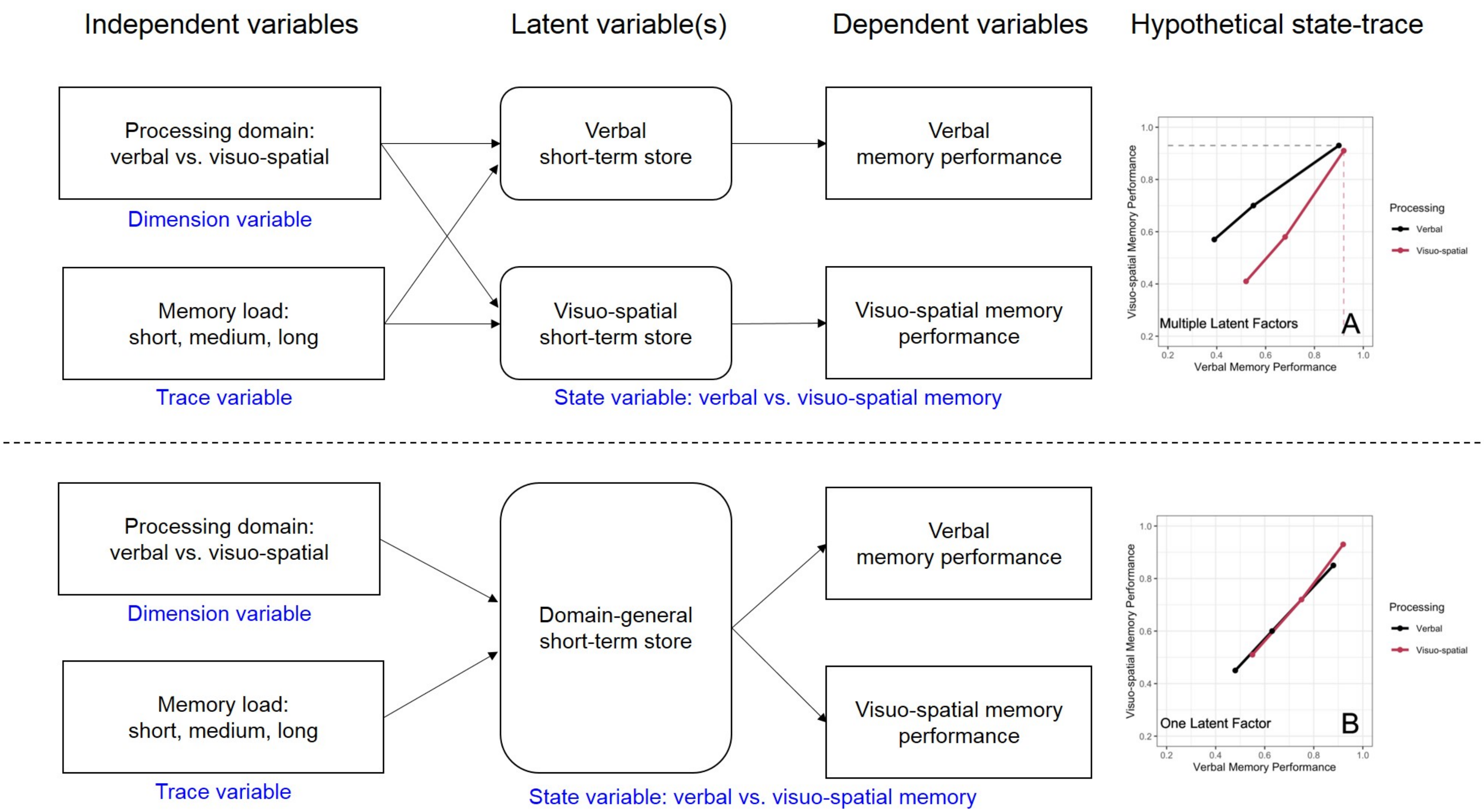

Figure 2

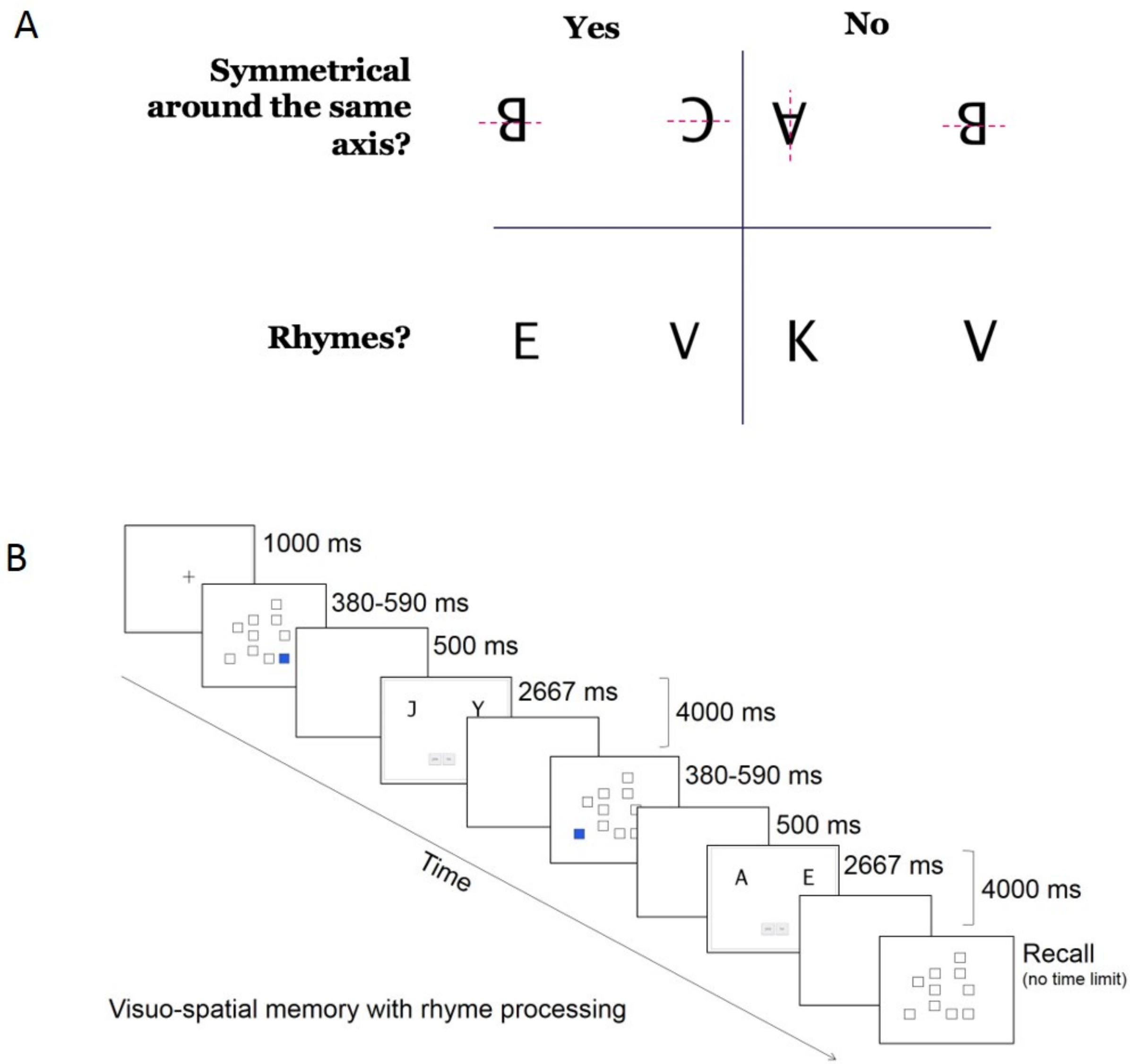
A
Yes
No
Symmetrical around the same axis?
Rhymes?
E
V
K
V
B
1000 ms
380-590 ms
500 ms
J
Y
2667 ms
4000 ms
380-590 ms
500 ms
A
E
2667 ms
4000 ms
Recall
(no time limit)
Time
Visuo-spatial memory with rhyme processing

Figure 3

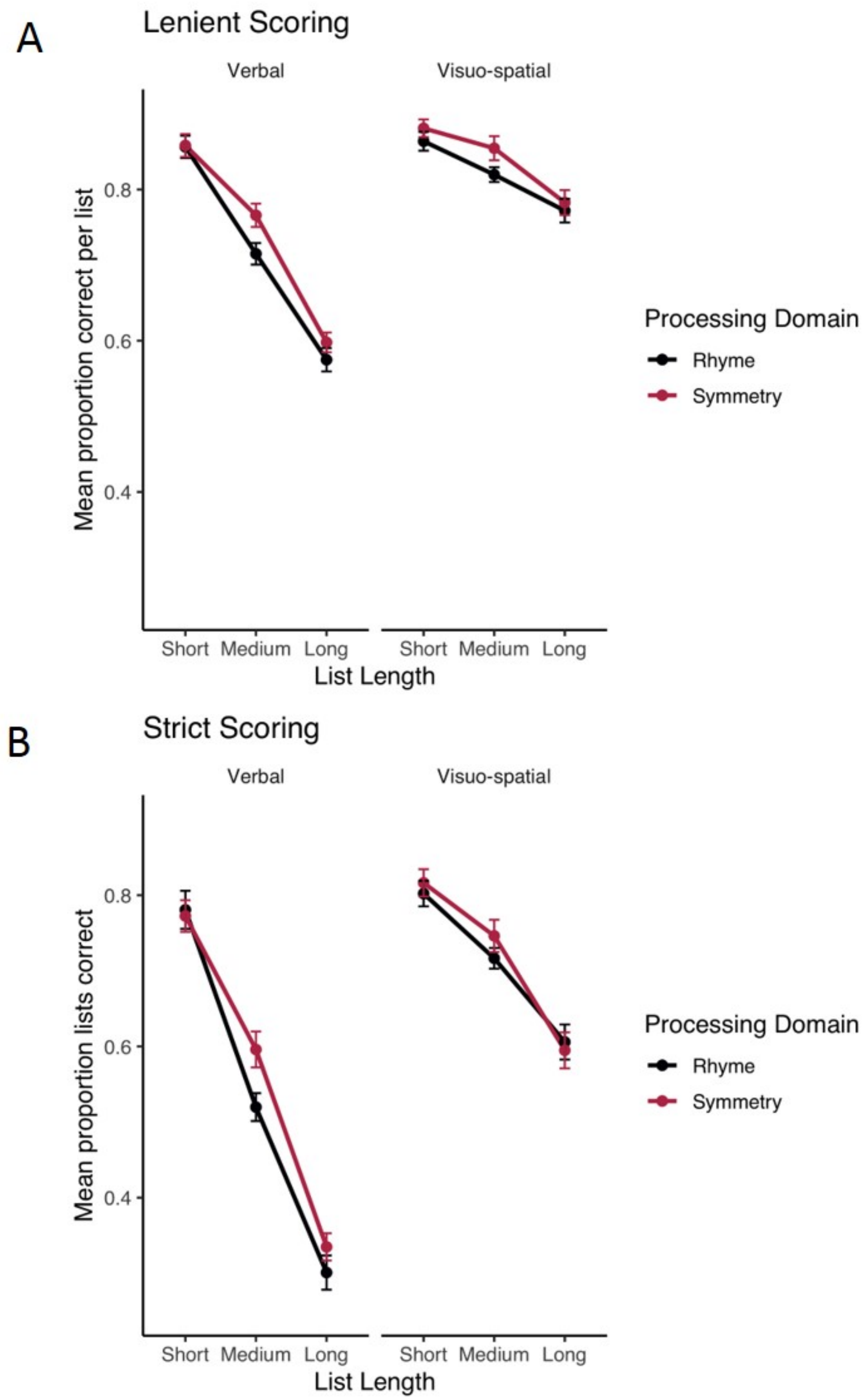
A
Lenient Scoring
Verbal
Visuo-spatial
Mean proportion correct per list
0.8
0.6
0.4
Short
Medium
Long
Short
Medium
Long
List Length
Processing Domain
Rhyme
Symmetry
B
Strict Scoring
Verbal
Visuo-spatial
Mean proportion lists correct
0.8
0.6
0.4
Short
Medium
Long
Short
Medium
Long
List Length
Processing Domain
Rhyme
Symmetry

Figure 4 – Panel A

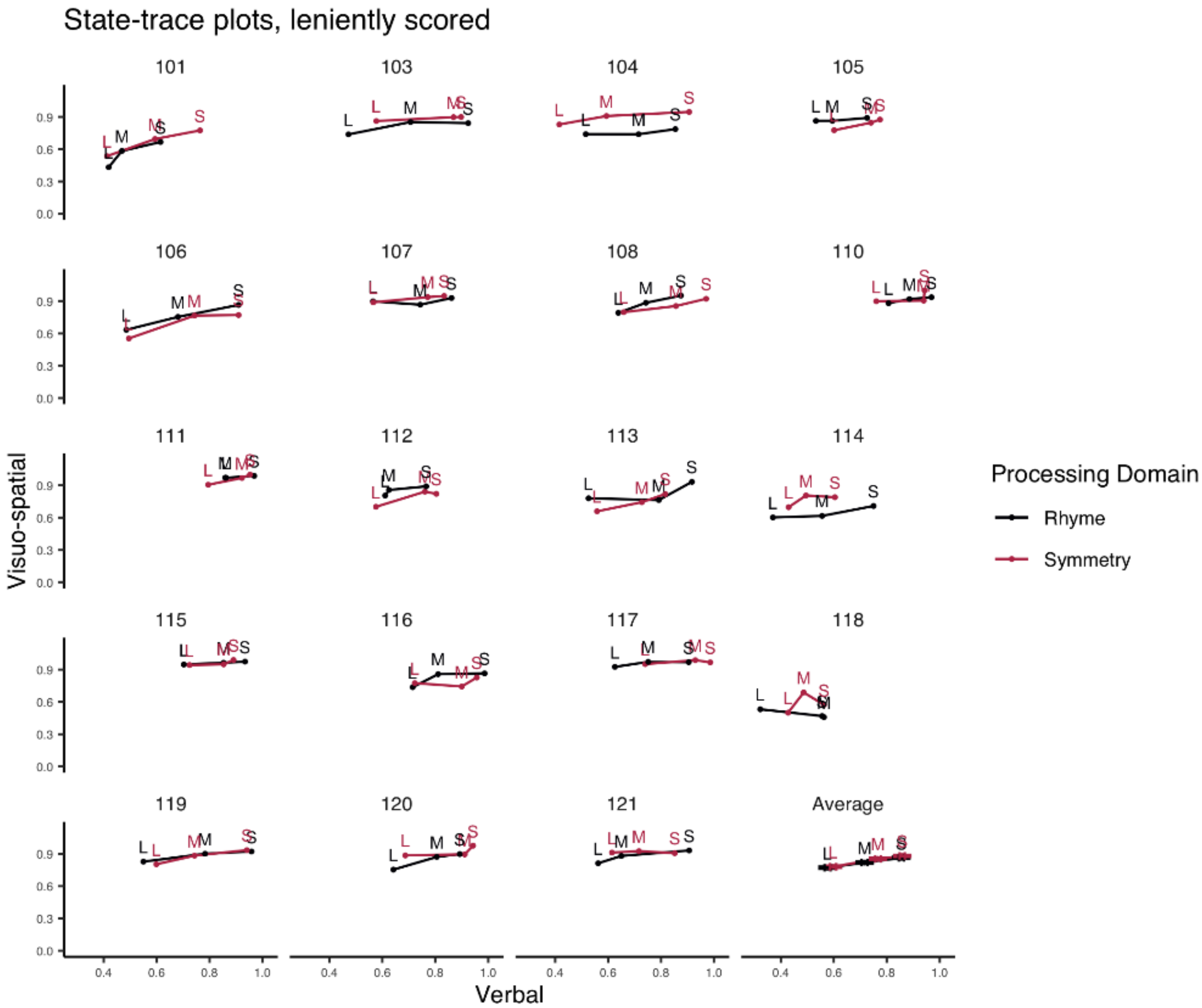

Figure 4 – Panel B

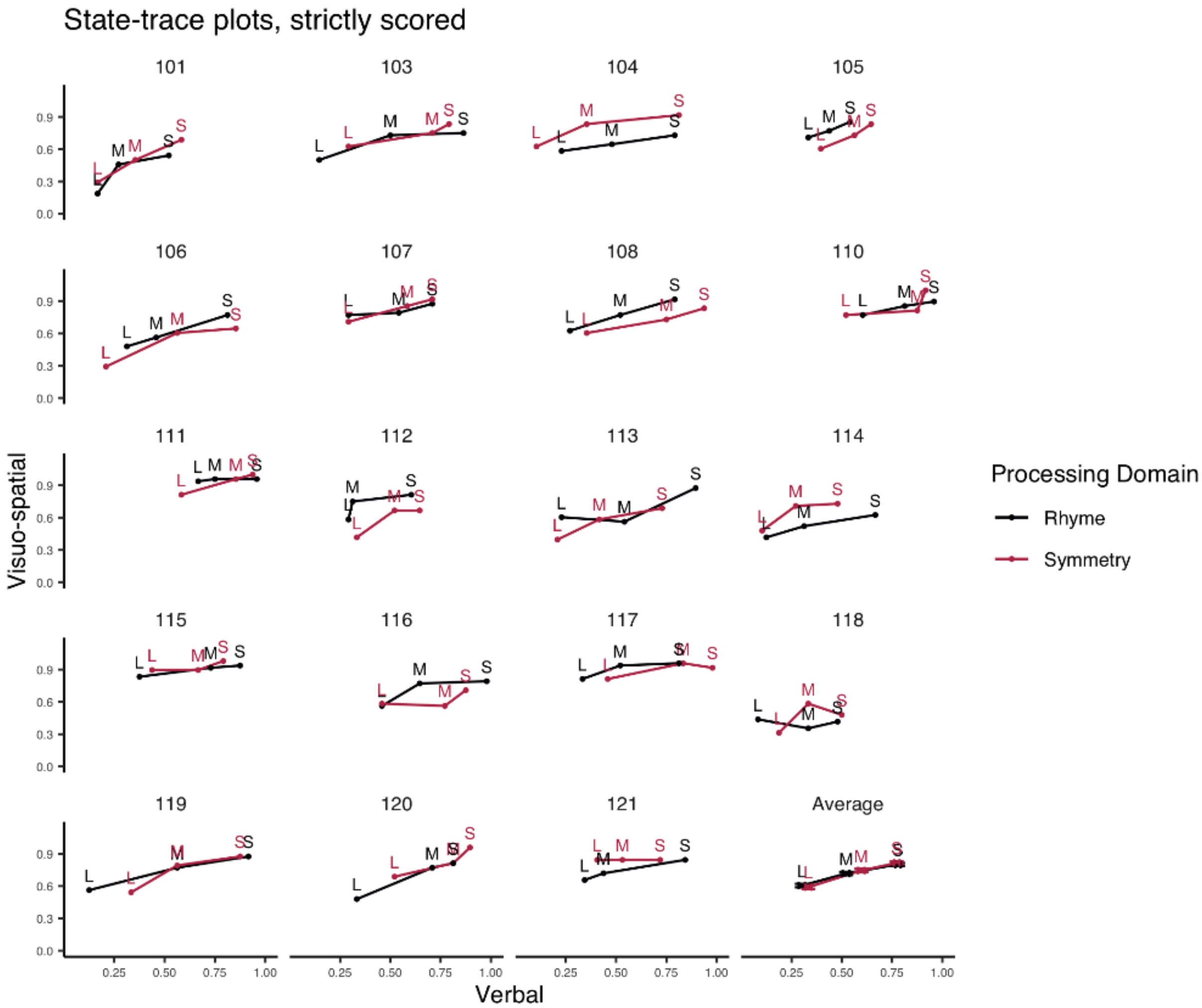

Figure 5

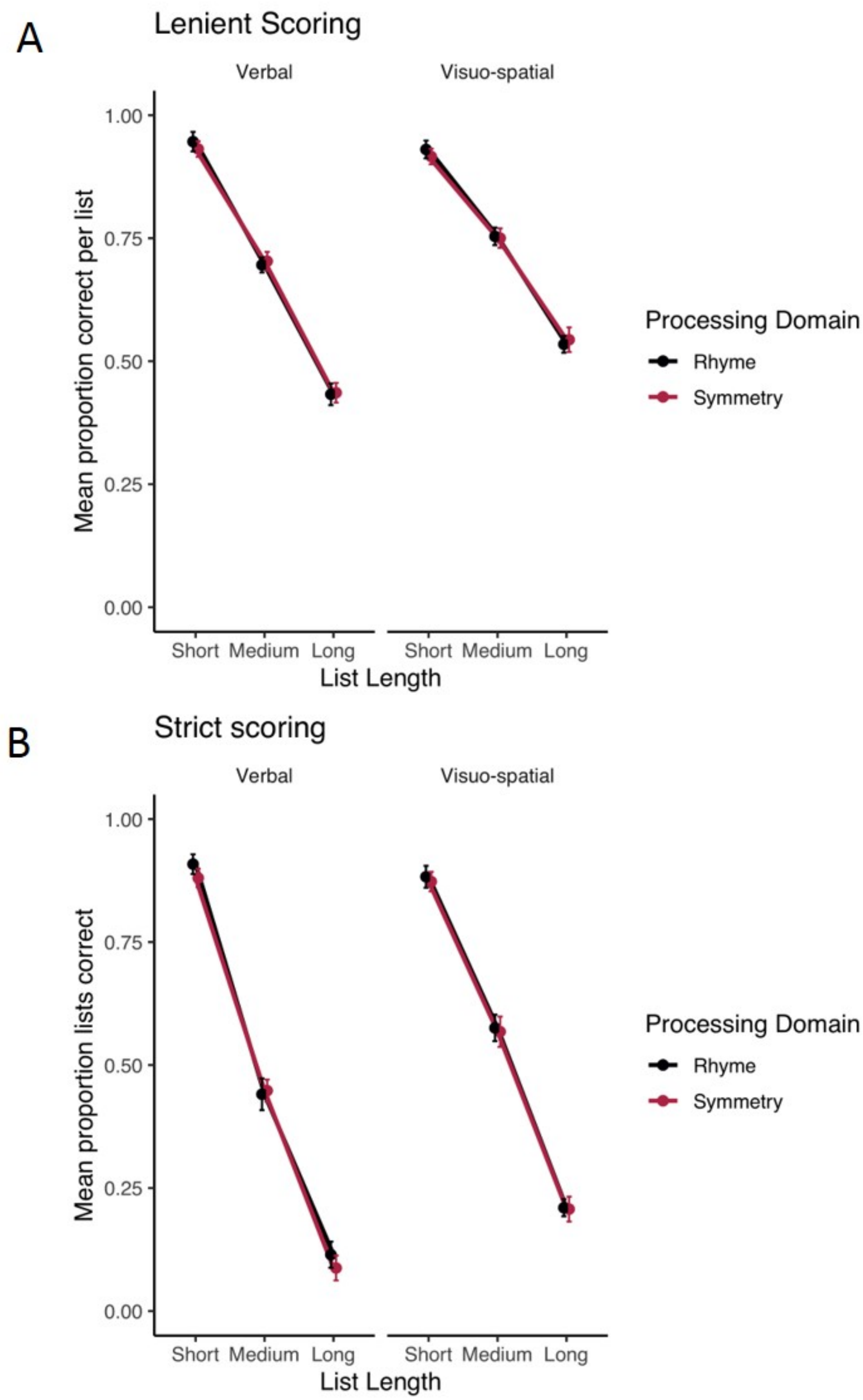
A
Lenient Scoring
Verbal
Visuo-spatial
1.00
0.75
0.50
0.25
0.00
Mean proportion correct per list
Short Medium Long
Short Medium Long
List Length
Processing Domain
Rhyme
Symmetry
B
Strict scoring
Verbal
Visuo-spatial
1.00
0.75
0.50
0.25
0.00
Mean proportion lists correct
Short Medium Long
Short Medium Long
List Length
Processing Domain
Rhyme
Symmetry

Figure 6 – Panel A

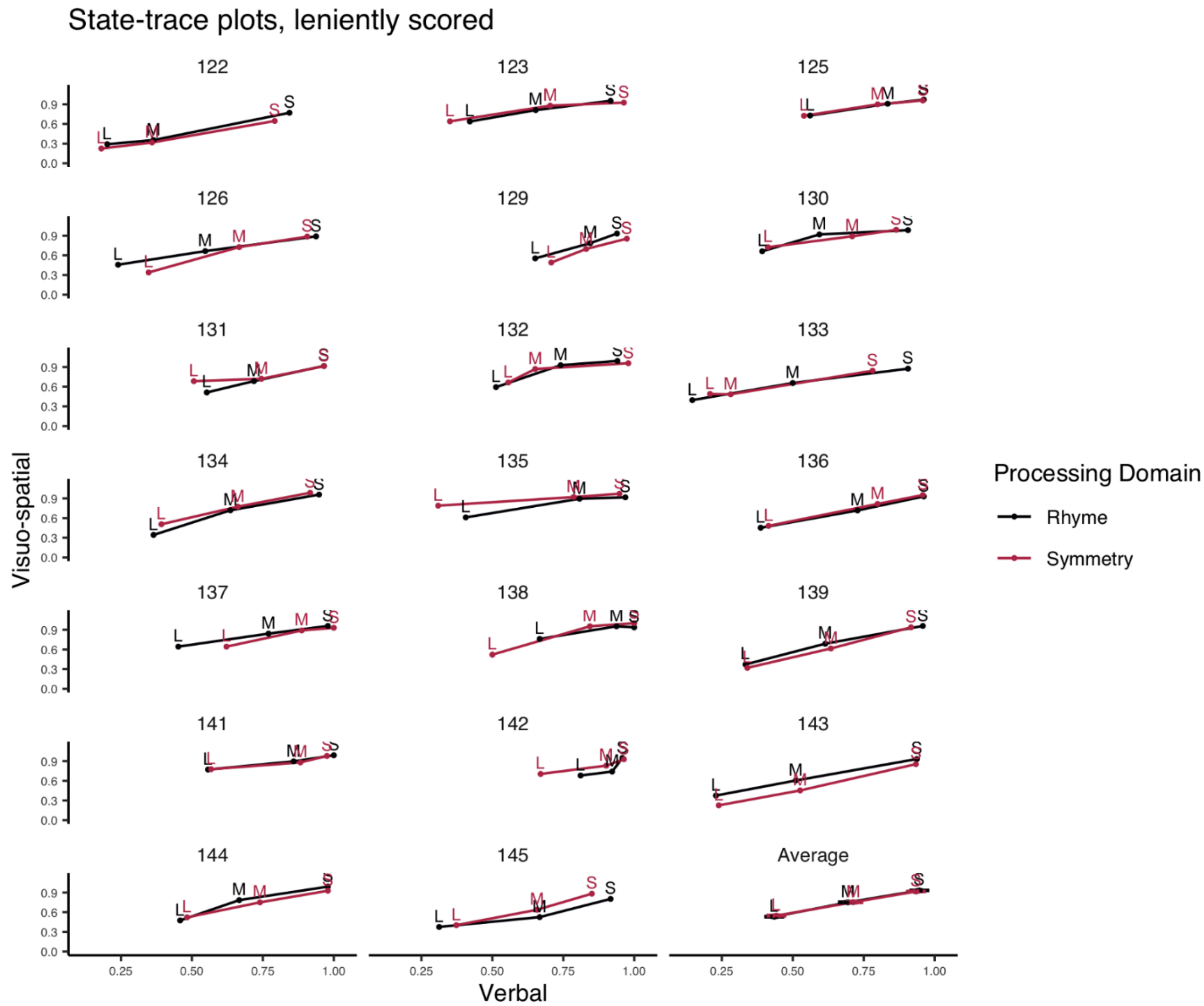

Figure 6 – Panel B