Journal of Memory and Language 108 (2019) 104032

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

A matter of priorities: High working memory enables (slightly) superior value-directed remembering

Michael L. Griffin*, Aaron S. Benjamin, Lili Sahakyan, Sarah E. Stanley

University of Illinois at Urbana-Champaign, United States

ARTICLE INFO

Keywords:
Working memory
Cued recall
Value-directed remembering

ABSTRACT

People with larger working memory capacity exhibit enhanced free recall. One explanation for this relationship is that the *strategies* that people bring to the task of learning and retrieving are superior in learners with high working memory. There is ample evidence that learners with high working memory do indeed bring better strategies to both encoding and retrieval, but as yet little evidence of whether higher working memory is related to greater effectiveness in prioritizing information across materials that differ in value. Using the *value-directed remembering* paradigm of Castel, Benjamin, Watkins, and Craik (2002), we examined whether learners with high working memory capacity show a particular advantage in remembering materials that are of high value. Across four experiments, we found that high working memory capacity led to a selective preference for remembering high-valued word pairs, but the effect was very modest and does not provide a complete picture of the relationship between working memory and recall.

Introduction

People with high working memory capacity show benefits on a wide variety of cognitive tasks, including measures of fluid intelligence like Raven's progressive matrices, the game of bridge, and SAT scores (Engle, Tuholski, Laughlin, & Conway, 1999; Clarkson-Smith & Hartley, 1990; Daneman & Carpenter, 1980). Of particular interest is the fact that they demonstrate superior recall (Unsworth, 2007; Unsworth, 2010)—a task that, on its face, places very different demands on the learner than do working memory tasks. One candidate for understanding the relationship between recall and working memory is strategy use. People may have a general capacity for managing the demands of memory tasks by the use of mental strategies that effectively retain current or critical information and discard outdated or unimportant information.

One important aspect of effectively confronting a memory task is *prioritization*. Students studying for a course always demand to know which material is "most important"—that is, most likely to appear on the test; presumably they would use this information, were they ever to receive it, in service of allocating encoding resources. In this paper, we examine the ability of those with high working memory to *prioritize* information they are studying for a later memory test. We do so by taking advantage of the *value-directed remembering* paradigm of Castel, Benjamin, Craik, and Watkins (2002), which provides explicit and clear values for each individual memorandum and provides a means of evaluating the selectivity of remembering. If the benefits of working memory are due to the superiority of the strategic processing that they bring to memory tasks, then one arena we should see these strategies at work is in focusing their study on the most relevant information.

Strategy use and working memory capacity

Two types of evidence suggest that the benefit enjoyed by learners with high working memory (HWM) capacity can be attributed to strategies that people bring to the task. In one literature, the imposition of specific strategies is shown to reduce or eliminate the benefit normally held by HWM learners. The implication of such results is that people with HWM spontaneously use effective strategies at a greater rate or more efficiently than do low working memory (LWM) people, and that the imposition of good strategies reduces the "strategy gap." In another literature, strategies are measured directly—usually by self-report—and the effect of those reported strategies on the relationship between WM and recall is assessed. Here we provide a brief review of evidence from each of these domains.

Strategy imposition

It is known that instructing subjects in a strategy for a working

* Corresponding author at: Psychology Department, University of Illinois at Urbana-Champaign, 603 E. Daniel Street, Champaign, IL 61820-6232, United States.
E-mail address: mgriffn2@illinois.edu (M.L. Griffin).

https://doi.org/10.1016/j.jml.2019.104032
Received 7 November 2018; Received in revised form 5 May 2019; Accepted 11 June 2019

memory task can improve performance. For example, McNamara and Scott (2001) showed that instruction on "chaining" techniques—linking to-be-remembered words into an ongoing story—increased performance on an operation span task. This result indicates that more is at work during memory-span tests than just inherent capacity limitations—strategies for effectively organizing the material can enhance performance, just like in traditional long-term memory tasks. Experiments that examine the gap in performance between HWM and LWM subjects have found that some instructions do not reduce this gap (Turley-Ames & Whitfield, 2003) but that others do (Unsworth, Brewer, & Spillers, 2013).

A relevant variable appears to be whether the instructed strategy is effective, and whether it can be applied to both the working memory task and the long-term memory task (Bailey, Dunlosky, & Kane, 2008). Turley-Ames and Whitfield (2003) instructed learners to use rote rehearsal, which is not a particularly effective strategy (Benjamin & Bjork, 2000), and examined long-term comprehension as the criterion variable. Unsworth et al., on the other hand, used highly effective retrieval cues (Tulving & Thomson, 1973).

Unsworth et al. (2013) reported a good example of the benefits of strategy imposition at the time of retrieval. In their task, subjects produced the names of animals from semantic memory. In the control condition, no constraints were imposed on this process. In the free-cue condition, subcategories of animals (e.g., pets and farm animals) were presented on the edge of the screen and remained for the duration of the task. In the forced-cue condition, participants were given one subcategory and spent the next 20 s recalling animals from only that category; then the cue switched and the next 20 s were devoted to the new cue; the process continued until all 15 categories were cycled through. Overall performance was highest in the forced-cue condition, followed by free-cue, and then control. Most importantly, working memory differences in the generation task were smaller with superior strategies: group differences were largest in the control condition and were nearly absent in the forced-cue condition.

This effect can also be seen in long-term memory tasks. In a directed forgetting study, Delaney and Sahakyan (2007, Exp. 2) showed that a difference in recall performance between HWM and LWM subjects can be reduced by applying a story mnemonic during encoding. Under normal conditions, WM predicts recall of the to-be-remembered words, but this effect was absent when participants were asked to construct a story using each of the words in the study list as they were presented. Interestingly, WM differences remained for to-be-forgotten information: participants with HWM were less able to recall the list they had been asked to forget than were participants with low working memory span (LWM). This result suggests that high working memory span also confers benefits in the degree to which outdated information can be discarded.

Strategy assessment

Other studies have revealed a variety of superior techniques utilized by learners with high working memory span. McNamara and Scott (2001) asked participants about strategies they used on an operation span test. They found that working memory span scores were higher for people who self-reported using normatively more effective strategies. Similarly, Bailey et al. (2008) found that self-reported effective strategies mediated part of the relationship between working memory span and episodic recall. They also proposed that other failures to find mediating relationships revealed that the learned strategies did not generalize across tasks, as might be expected for reading comprehension (cf. Turley-Ames & Whitfield, 2003). Such results are to be expected because there are more strategies that benefit two forms of recall (e.g., elaborative rehearsal or chaining), but fewer that enhance both recall and comprehension, the latter of which benefits from a distinct set of behaviors, like skimming or summarizing main ideas.

Strategies were assessed more directly in the animal-listing task discussed earlier (Unsworth et al., 2013). Those with high working memory retrieved more unique animals, but also seemed to retrieve them differently: HWM subjects tapped more different clusters (such as house pets or farm animals), and retrieved a greater proportion of their total recall from those clusters. The use of this strategy was borne out by direct questioning: when asked how they attempted to retrieve animals, HWM participants more often reported using a 'general to specific' strategy, where they began with a general type/category of animal (e.g., bird), and then listed exemplars, while low working memory participants were more likely to report using no organizing strategy.

Another recent study by Unsworth has tracked WM and strategy use in a more conventional free recall task (Unsworth, 2016). Participants studied words that were either self-paced, or presented for either 1 s or 4 s. After a short distractor task, participants had 1 min to recall the studied words, and then self-reported their encoding strategies. WM was expected to correlate with strategy use, particularly in the longer study-time condition and the self-paced condition, where more elaborate encoding strategies could be brought to bear more easily. Their results indicated that strategies during study did indeed affect recall. Self-reported strategies were split into 'effective' and 'less effective' groups on the basis of known effects from the literature—effective strategies included imagery, sentence generation, and grouping; less effective strategies included passive reading and simple repetition. Effective strategies were more commonly reported with greater study time. But strategy use also varied with WM—HWM subjects reported the use of more effective strategies in all conditions. Interestingly, they found that WM has only an indirect effect on recall, being mediated through intrusions, inter-response times, and reported strategy use.

Limitations of extant research and benefits of value-directed remembering

Although the imposition of a strategy can yield impressive benefits and even close the gap between HWM and LWM learners, the results do not directly tell us about the types or effectiveness of strategies that individuals use in the absence of direct instruction. Self-reports are a partial solution, but have a different set of costs. Self-reports rely on introspection, which risks being incomplete, biased, or based heavily on inference due to failures of memory when collected after the fact (Ericsson & Simon, 1993). If strategies are assessed during encoding, the reports themselves are more accurate, but the solicitation may lead subjects to change strategies (cf., Mitchum, Kelley, & Fox, 2016).

Our examination of how well subjects can prioritize information in memory avoids some of these traditional difficulties. Like remembering in the real world, studies of prioritization force the learner to make decisions about what to ignore and what to emphasize. In metacognition, prioritization is often framed as a choice of what to (re)study. For instance, time pressure leads learners to shift their study emphasis from difficult to easy words, to ensure at least a low level of successful recall (Son & Metcalfe, 2000). Learners can also capably prioritize even when values are not explicitly provided. In self-guided learning, for example, participants choose a subset of items for restudy, and honoring this choice leads to better memory than dishonoring it (Kornell & Metcalfe, 2006; Tullis, Fiechter, & Benjamin, 2019).

We can evaluate the ability to prioritize directly with the value-directed remembering paradigm. In the task, a word or word pair is assigned a point value, and participants are told to try and maximize <u>not</u> the number of words remembered, but the <u>point total</u>, encouraging a prioritization or selectivity in encoding. An early study using value-directed remembering (Castel et al., 2002) had subjects learn and then immediately recall multiple lists of 12 words each, with each word within a list having a unique point value (from 1 to 12). Selectivity was scored by examining the degree to which the recalled items selectively included high-value words.

Here we evaluate whether those with higher working memory also exhibit greater selectivity. If there are individual differences in strategy use that partially mediate the relationship between working memory

and higher order cognition (including recall), then those with higher working memory should not simply remember more words, but should also show better priorities in choosing which words to remember. In that view, those with high working memory are generally more strategically savvy at encoding, regardless of task. If the goal is free recall, they will employ effective strategies like generation or imagery (as in Unsworth, 2016); in value-directed remembering, they will hone in rapidly on the effective strategy of allocating more attention, effort, and rehearsal time to items of higher priority.

However, if this partial mediation depends on a common strategy being used in both tasks—for example, if imagery helps you remember items in both the span task and in free recall, then prioritization will fail to mediate the relationship if cannot be profitably used across task.

In fact, the extent to which prioritization should be expected to predict overall memory is complicated.

There are examples in the literature of both increases and decreases in selectivity with variables that decrease memory overall. Older adults with mild forms of Alzheimer's disease exhibited both poorer memory and lower selectivity than matched older control subjects (Castel, Balota, & McCabe, 2009). Yet older adults with no memory pathology show poorer memory but higher selectivity than younger adults (Castel et al., 2002; Castel et al., 2011). One recent paper examined the effects of working memory on prioritization (Robison & Unsworth, 2017). They used a variant of free recall that afforded participants considerable explicit control over which words to study, and additionally instructed participants on what study strategy to employ. They found that working memory was related to prioritization in this task and also that, with certain study strategies, differences in prioritization as a function of working memory were eliminated. The results from their work differ in important ways from our own, so we reserve a more detailed examination of their procedure and results for the General Discussion.

For present purposes, we used a procedure similar to that used by Castel et al. (2002). The most important change is the use of a cued, rather than free, recall task. We made this choice so that we could better focus on strategies specific to encoding (as opposed to retrieval). In contrast, the strategy that Robison and Unsworth provided to participants instructed them on how to *study*, but participants were free to employ (potentially differing) strategies at time of retrieval. On tests of free recall, people may selectively output the highest valued words first, in order to avoid output interference for those critical items. This choice would lead to a higher selectivity score, but involves an element of strategizing at the time of test—thus making it harder to tell whether prioritization was also happening at encoding. With cued recall, we controlled when each item was tested. Remaining changes to the Castel et al. procedure were minor and mostly procedural in nature; they are indicated in the Methods sections.

In all the experiments reported here, each participant completed three span tasks (adapted from Unsworth, Heitz, Schrock, & Engle, 2005; Unsworth, Redick, Heitz, Broadway, & Engle, 2009), and then participated in the value-directed remembering task. Before outlining the individual experiments, we describe the involved tasks. We developed our own implementation for these tasks in Adobe Flash so that they could be administered online.

Operation span

In the operation span task, participants solved arithmetic problems while trying to remember letters that appeared between those problems. All letters were from the pool (F,H,J,K,L,N,P,Q,R,S,T,Y), with 3–7 letters being presented in a given block. Following 3 blocks of practice, there were 15 blocks (3 of each length, in random order) that were scored. Arithmetic problems were of the form $(4 * 3) - 7 = ?$ The first operation was always multiplication/division, and the second was always addition/subtraction. Once they had the answer, participants clicked to bring up a new number with YES and NO below it – they clicked YES if the new number was the correct answer to the equation, and NO if it was not. After answering YES or NO, a letter appeared for 800 ms, and was followed immediately by the next math problem. If participants took too long on the math equation, a red *TOO SLOW!* message replaced the equation for 1 s, and then the letter would appear. Participants were instructed to keep their performance on the math above 85%, while remembering as many letters as possible. Between blocks, a running average of math performance was shown (presented in green if at or above 85% and in red if below), and subjects were prompted to work more carefully on the equations if their performance dropped below 85%.

For practice, participants began with 15 trials of only the math problems, with their performance on the practice setting an upper limit on the time they could spend on each math problem in the rest of the task. They also had 3 short (2, 2, and 3 letter) blocks of practice on the actual task, before beginning the longer sets they would be evaluated on. The task was scored as the proportion of letters recalled in the correct serial position within a block. Thus, the maximum score for the 15 blocks was 15, with errors in the more difficult blocks decreasing one's total score less.

Reading span

In the reading span task, participants were given sentences such as *Cows are four-legged animals that can quickly climb trees* and asked to judge them as either true or false, all while trying to remember letters that appeared between each sentence. For each block, there were 2–6 letters to remember, with 10 blocks in total (2 of each size). Each trial played out similarly to operation span: A sentence appeared first, participants clicked once after they had read the sentence, and on the next screen clicked YES if the sentence was true or NO if the sentence was false. Then a letter appeared for 800 ms, with the next sentence directly following it. Letters were drawn from the same common pool as the operation span task.

At the start, participants practiced the sentence task on its own for 15 trials, with their performance on the practice setting an upper limit for the time they could spend on each sentence during the actual task. They also had 2 short practice blocks of the full task with each having only 2 letters to remember, before beginning the scored portion of the task. Participants were told to keep performance on the sentence task above 85% while trying to remember as many letters as possible, with a running average of their sentence verity performance again being displayed between blocks. The task was scored as proportion of letters recalled in the correct serial position within a block, summed across blocks.

Symmetry span

Participants judged whether grids of black and white squares were vertically symmetric while trying to remember the location of red squares on (smaller) 4×4 grids that appeared between judgments. Each block had 2–5 symmetry judgments and red squares to remember, with 12 blocks in total, 3 for each length. At the end of each block, participants clicked on a blank grid to show where each red square had been, and order mattered: if the third red square was in the upper left, an upper left location would be accepted as correct only if it was the third location entered. Red square locations did not repeat within a block. The symmetry judgments were done by mouse: people clicked once they determined whether the image was symmetric or not, and then clicked either YES or NO on the next screen to indicate their response. After the response, the smaller grid with one red square appeared for 650 ms, and was followed by a blank screen for 550 ms before the next symmetry judgment began.

As in the other two span tasks, participants practiced on its components prior to the full task. There were 2 blocks of practice (length: 2 and 3) for the square task, and 6 trials of practice for the symmetry task. They also received two blocks of practice on the full task (length: 2 and

3), before beginning the full task. Their speed on these last two blocks were again used to create an upper limit (2.5 SDs above their average) on the time they could spend for the symmetry portion of the span task, and as before participants were told to remain at least 85% accurate on the symmetry portion while remembering as many red square locations as possible. Scoring was again determined from the proportion of items recalled in the correct serial order within a block, summed across blocks.

Value-Directed remembering

Participants studied one list of 54 word-pairs with the instructions to maximize the point total associated with the items that they successfully recalled. Each word pair's presentation lasted for 6 s in total. First, the point value was shown alone for 1 s, then the word pair appeared just above that value for 4 s; both were replaced by a blank screen that lasted for 1 s. This process continued for all 54 word pairs.

The words were chosen from the University of Florida free association word norms (Nelson, McEvoy, & Schreiber, 1998). They were 3 to 8 letters long, with low but non-zero cue to target strength (FSG between .01 and .012), and no reported target-to-cue strength. Example pairs include *barrel – drum, stay – home,* and *collar – blouse*.

At test, subjects were presented with the left (cue) word, and told to type in the right (target) word that had been previously paired with it. There was no time limit for this portion of the task, and participants were told that they could either guess or type 'SKIP' if they didn't remember the word that had been paired with the cue.

Experiments 1a and 1b

In the first set of experiments reported here, subjects studied a single list of word pairs (and point values) and experienced only a single test. The two experiments constitute near-identical replications, with one being performed in a university laboratory using college students and the other using an online sample.

Method

Subjects

Experiment 1a was run online on Amazon's Mechanical Turk. On Mechanical Turk, workers were paid $2.50 for finishing the study, and were only eligible to participate if they were in the U.S., had completed at least 5,000 HITs, and had at least a 98% approval rating. 106 subjects participated, with five lost due to technical problems, leading to 101 subjects whose data were used in the analyses. For Experiment 1b, 97 University of Illinois students participated, with two lost to technical problems, leading to 95 subjects whose data were used in the analyses.

Materials

In both experiments, each participant completed the three span tasks, and then participated in the value-directed remembering task. The order of the working memory span tasks was random for each subject. In Experiment 1a, participants were run online on Mechanical Turk. In Experiment 1b, participants were University of Illinois students, and were run individually in the Human Memory and Cognition laboratory.

Experiments 1a and 1b also differed in the range of point values assigned to word pairs in the value-directed remembering task. In Experiment 1a, the value for each pair ranged from 0 to 5 points. In Experiment 1b, this was simplified to three possible point values: 0, 5, or 10 points. In both experiments, each possible point value was equally represented. With 54 word pairs in total, this meant there were nine word-pairs tied to a given point value in Experiment 1a, and eighteen tied to each value in Experiment 1b. Because the ultimate measure of selectivity uses each individual's set of possible point values to compute the optimality of a subject's response, this minor difference in scales is accounted for in a straightforward way.

Procedure

Prior to beginning the experiment, all subjects either filled out an online consent (if on Mechanical Turk), or completed a consent and demographics form (if run in the lab), and then were instructed about the tasks. As described above, each WM task included a short practice on the individual components of the task before the combined task began. After completing all three WM tasks, they moved on to the cued recall portion of the experiment.

Results

Prioritization in recall was scored as the proportion of points achieved relative to the maximum number of points available, conditional upon the total number of items recalled (see also Castel et al., 2002; Watkins & Bloom, 1999):

$$Selectivity = \frac{Obtained\ points - Chance\ points}{Maximum\ points - Chance\ points}$$

Chance points is the product of number of items recalled and the expected value of an item randomly recalled without respect to point value (2.5 in Experiment 1a, 5 in Experiment 1b). The selectivity score ranges from −1 to 1, with 1 being perfect selectivity, and 0 indicating no selectivity.

For all analyses, a single composite score for WM was calculated via confirmatory factor analysis, with a model fit to the data defining a single latent 'overall' WM variable, with the three span scores loading onto it. Once fit, the resulting score for each subject was used as the composite measure of WM capacity. Although the primary analyses all treat WM as a continuous variable, some figures split WM into 'high' and 'low' groups via a median split for illustrative purposes.

Experiments 1a and 1b employed the same procedure, with only a difference in the range of point values assigned to the memoranda, and are analyzed together here. The primary analyses of interest were the correlations between working memory, recall accuracy, and prioritization. These combined data are also analyzed using Bayesian analyses; in particular, Bayes factors (in favor of the alternative) are calculated for our correlations of interest. The principal advantage of Bayesian analysis is that it can evaluate evidence for and conclude in favor of the null hypothesis, whereas traditional null hypothesis tests can only fail to reject the null hypothesis. Given that many of the correlations between working memory and long-term memory are small in magnitude (e.g., Unsworth, 2010), it is valuable to be able to distinguish between small but present correlations, and correlations that are indistinguishable from 0.

Bayesian analyses require specification of prior distributions for null and alternative hypotheses. Our analyses follow Rouder and colleagues' (2009) recommendation and use a Jeffreys-Zellner-Siow (JZS) prior. The goal of this prior is to be 'objective', with minimal assumptions about the range of possible effect sizes; it follows a Cauchy distributed range of standardized effect sizes. Lastly, while there are no critical values for Bayes factors, Jeffreys (1961) provides a set of guidelines that has come into common use: a BF > 3 indicates *some evidence*, > 10 indicates *strong evidence*, and > 30 is *very strong evidence*.

Experiment 1a

Across subjects, mean recall accuracy was .312 (SE = .025), and mean selectivity was .154 (SE = .036). Table 1 provides the full set of descriptive statistics for these measures across each experiment. Fig. 1 shows the relationship between working memory and recall, as well as between working memory and selectivity. Fig. 2 shows recall accuracy for word-pairs at each point value. Working memory correlated with

Table 1
Descriptive Statistics for recall accuracy and selectivity across each experiment.

Descriptive statistics for accuracy and selectivity		Mean	Variance	Skew	Kurtosis	Min	Max	Reliability
Accuracy								
	Experiment 1a	0.31	0.06	0.58	−0.47	0.00	0.98	0.95
	Experiment 1b	0.35	0.03	0.26	−0.63	0.02	0.76	0.88
	Experiment 2a	0.32	0.04	0.78	0.77	0	1	0.96
	Experiment 2b	0.26	0.02	0.43	−0.47	0.01	0.69	0.92
Selectivity								
	Experiment 1a	0.15	0.13	−0.05	0.97	−1	1	0.50
	Experiment 1b	0.27	0.09	−0.69	2.21	−1	1	0.53
	Experiment 2a	0.52	0.11	−0.90	1.34	−0.75	1	0.81
	Experiment 2b	0.46	0.07	−0.63	0.16	−0.33	1	0.61

recall ($r = .342$, $t(98) = 3.603$, $p < .001$), with higher working memory predicting better accuracy on the cued recall test. Working memory also correlated with selectivity ($r = .228$, $t(90) = 2.222$, $p < .05$), with high span subjects better prioritizing high value word pairs (see Figs. 1 and 2).

Experiment 1b

Across subjects, mean recall accuracy was .346 (SE = .018), and mean selectivity was .273 (SE = .031). Working memory again reliably correlated with recall, $r = .225$, $t(93) = 2.223$, $p < .05$, but working

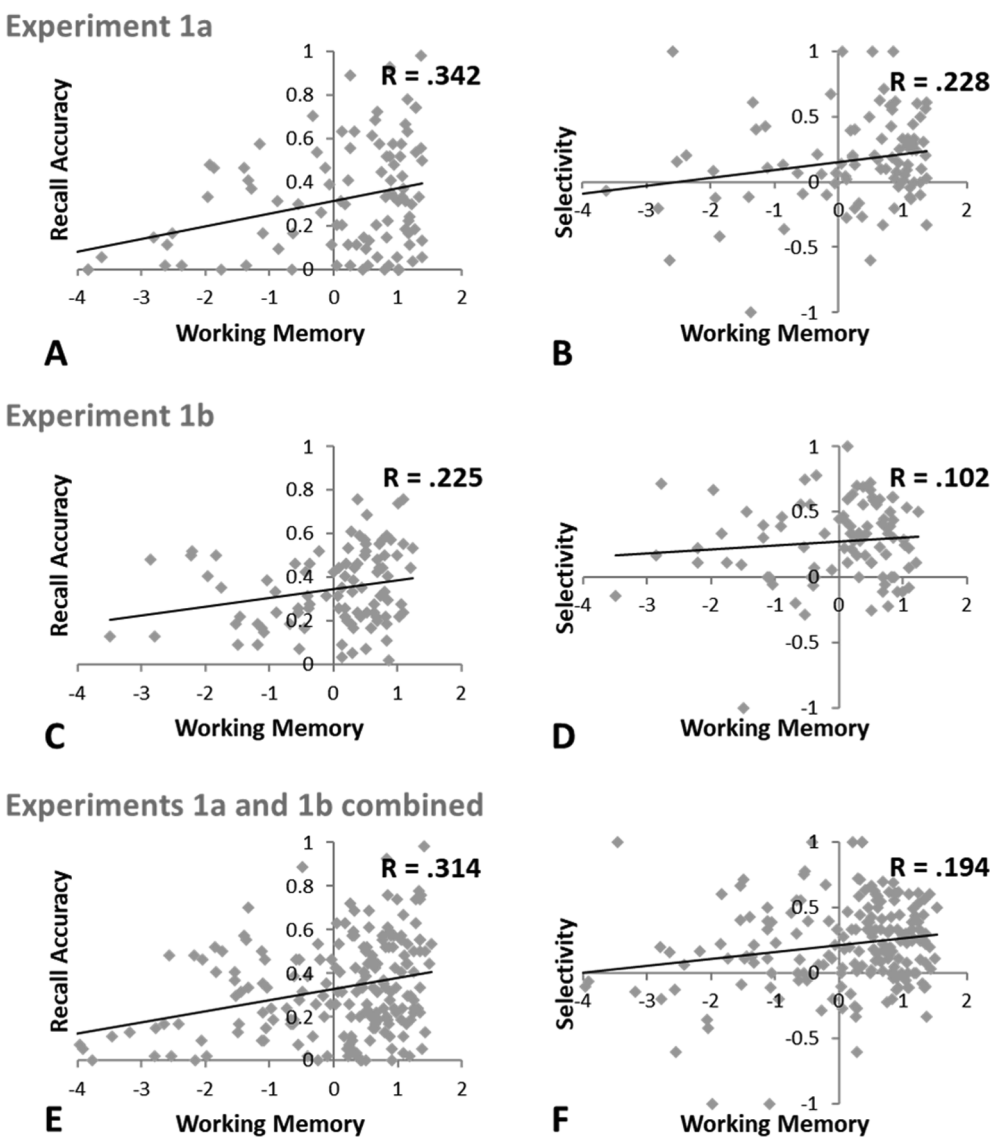

Fig. 1. The relationship between working memory and cued recall accuracy (left panels), and between working memory and selectivity (right panels) in Experiments 1a and 1b.

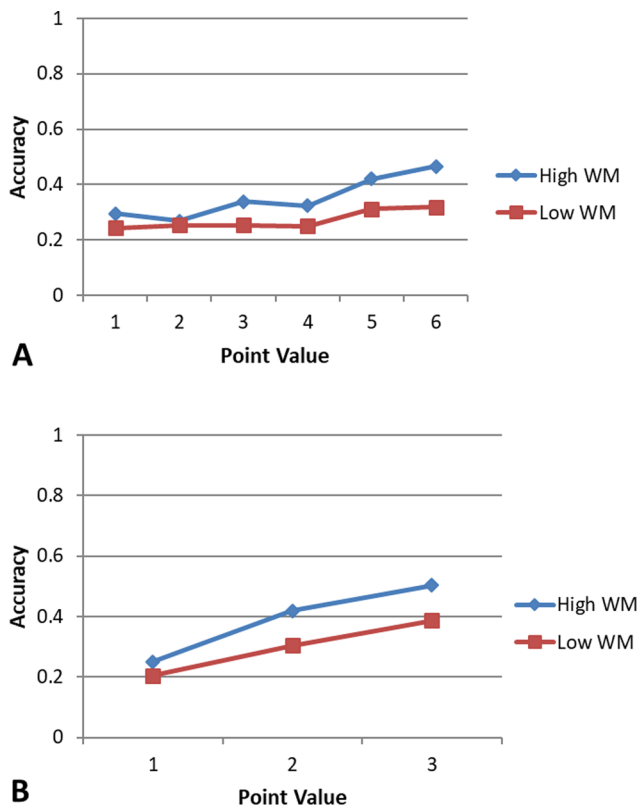

Fig. 2. Accuracy at each point value for those with high and low working memory in Experiments 1a and 1b.

memory did not reliably correlate with selectivity in this replication, r = .102, t(93) = 0.991, p = .32.

Experiments 1a and 1b combined

Consistent with prior work, working memory correlated with recall (r = .314, t(193) = 4.592, p < .001, BF_{10} = 1548.154), with higher working memory scores predicting greater accuracy on the cued recall test. Working memory also correlated with selectivity (r = .194, t(185) = 2.688, p < .01, BF_{10} = 2.68)—though more modestly—such that higher-span subjects were better able to prioritize higher value words.

Discussion

We found that working memory predicted recall, and also that participants prioritized high-value over low-value items in long-term memory. Both of these results are consistent with claims that learners are strategic in their approach to memory tasks (Unsworth, 2010; Castel et al., 2002). However, there was only weak evidence that working memory predicted prioritization: only one experiment yielded a correlation that reached significance, and in the combined analysis a Bayes factor indicates that the alternative is favored over the null hypothesis, but not convincingly. There was no apparent relationship between recall and selectivity, though the absence of that effect may reflect two offsetting factors. One can become more selective by attending selectively to high-value items, thereby decreasing total recall. But, in the population, people who recall more are probably also higher in selectivity. Consequently, it is hard to interpret the meaning of that effect's absence.

In Experiments 2a and 2b, we used a slightly different procedure that may allow easier detection of the relationship between working memory and prioritization. The first and most important change involved the number of study-test cycles: rather than study and test on a single long list of word pairs, there were 5 shorter lists, with a test following each one. Having the opportunity to experience the effects of one's encoding strategies on an actual memory test often leads learners to a better assessment of the effect of those strategies on memory (e.g., Benjamin, 2003; Finley & Benjamin, 2012; Sahakyan, Delaney, & Kelley, 2004; Tullis, Finley, & Benjamin, 2013; Tullis & Benjamin, 2012). To ensure that the task remained sufficiently difficult, even with these shorter lists, presentation time for each word pair was shortened from 6 s to 4 s. In this experiment, feedback on performance was provided after each list, giving the subjects the opportunity to shift to a more selective strategy over lists.

Experiments 2a and 2b

Method

Subjects

Experiment 2a was again run on Mechanical Turk. As in Experiment 1a, workers were only eligible to participate if they were in the U.S., had completed at least 5,000 HITs, and had at least a 98% approval rating, and were paid $2.50 for participating. Of the 102 subjects who participated, two experienced technical issues, leading to 100 subjects whose data were used in the analyses. For Experiment 2b, the in-lab replication, 97 University of Illinois students participated and all contributed data to the analysis.

Materials

The same working memory span tasks were used as in Experiments 1a and 1b. The value-directed remembering task was now composed of five study-test cycles, rather than one. Point values remained the same as Experiment 1b, with word pairs worth either 0, 5, or 10 points. Each study list was 21 word pairs long. The duration of a study trial was 4 s for each word pair. Each trial consisted of a 0.5 s presentation of the points alone, followed by 2.5 s with both points and the word pair present, and then a 1 s blank screen before the next trial began. A test followed each list. As before, participants could skip items during the test, and when each test phase was completed, a new screen appeared telling them how many points they had earned, and how much better or worse they had done compared to their previous study list.

In addition, a new set of stimuli were chosen. The word pairs in Experiment 2 were semantically unrelated, thus encouraging more idiosyncratic (and more variable) encoding decisions. Words that were 3–7 letters, with word frequencies between 5 and 400, were collected from the free association word norm database (Nelson et al., 1998), and shuffled into random pairs. Any pairs that by chance did have a recorded cue-to-target associative strength were eliminated, with 105 of the remaining word pairs used as the stimuli for these experiments.

Experiments 2a and 2b

As before, the only difference between Experiments 2a and 2b was the subject sample. Experiment 2a included paid subjects from Amazon Mechanical Turk; Experiment 2b included undergraduates from the University of Illinois who participated for course credit. The span tasks and value-directed remembering task were identical across the experiments, and both followed the general procedure below.

Procedure

The general procedure was largely unchanged from Experiment 1. Participants again completed symmetry, operation, and reading span tasks in a random order before moving onto the value-directed remembering portion of the experiment. For each of the span tasks, there was again a short practice session followed by the full task, and participants were encouraged to keep their performance on the intervening task above 85%. After all working memory tests were finished, they completed the five cued recall study-test cycles in value-directed remembering, and after each of the five were informed of their score on that cycle and how it compared to their previous score.

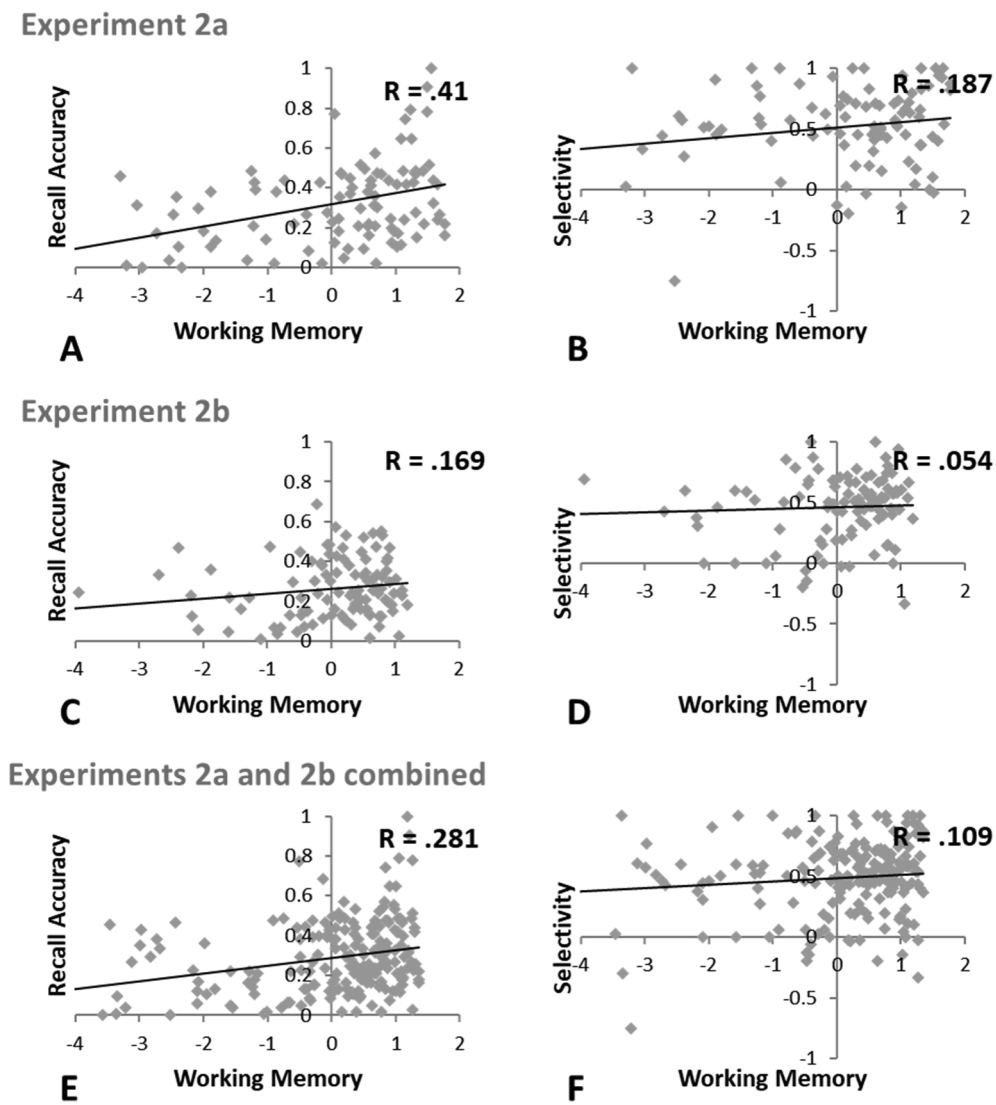

Fig. 3. The relationship between working memory and cued recall accuracy (left panels) and between working memory and selectivity (right panels) in Experiments 2a and 2b.

Results

Experiment 2a

Across subjects, mean recall accuracy was .317 (SE = .021), and mean selectivity was .519 (SE = .034). Fig. 3 shows the relationship between working memory and recall, as well as between working memory and selectivity. Fig. 4 shows the recall accuracy for low, medium, and high value word-pairs. In Experiment 2a, working memory reliably correlated with recall, r = .41, t(97) = 4.43, p < .001. Working memory marginally correlated with selectivity, r = .187, t(94) = 1.844, p = .068.

Experiment 2b

Across subjects, mean recall accuracy was .263 (SE = .014), and mean selectivity was .461 (SE = .026). Compared to Experiment 2a, the correlation between WM and recall was somewhat smaller, r = .169, t(104) = 1.752, p = .08. The correlation between WM and selectivity was barely in evidence, r = .054, t(104) = .554, p = .58.

Experiments 2a and 2b

Across Experiments 2a and 2b, working memory correlated with recall (r = .281, t(203) = 4.179, p < .001, BF_{10} = 304.575). Unlike Experiment 1, the correlation between working memory and selectivity was not reliable (r = .109, t(200) = 1.544, p = .12, BF_{10} = .253), so that higher working memory did not predict greater prioritization.

Although working memory did not reliably predict better prioritization, experience at the task did. Participants improved across lists: there was a main effect of list number on accuracy (F(4,812) = 35.714, p < .001), and also on selectivity (F(4,668) = 14.394, p < .001). Together, these results show that subjects could learn not just how to remember more, but how to become more efficient, over lists. Fig. 5 shows a breakdown of accuracy for each point value across lists, showing that enhancements in memory on later lists are localized to the higher-valued items.

Combined analysis

Table 2 shows how the working memory subtests correlated with each other, collapsed over all of the experiments[1]. Across all

[1] Results were similar between online studies and their in-lab replications. Experiment 1 had the larger departures, with reading span correlations being ~.1 below the omnibus analysis in 1b. For experiments 2a and 2b, subtest

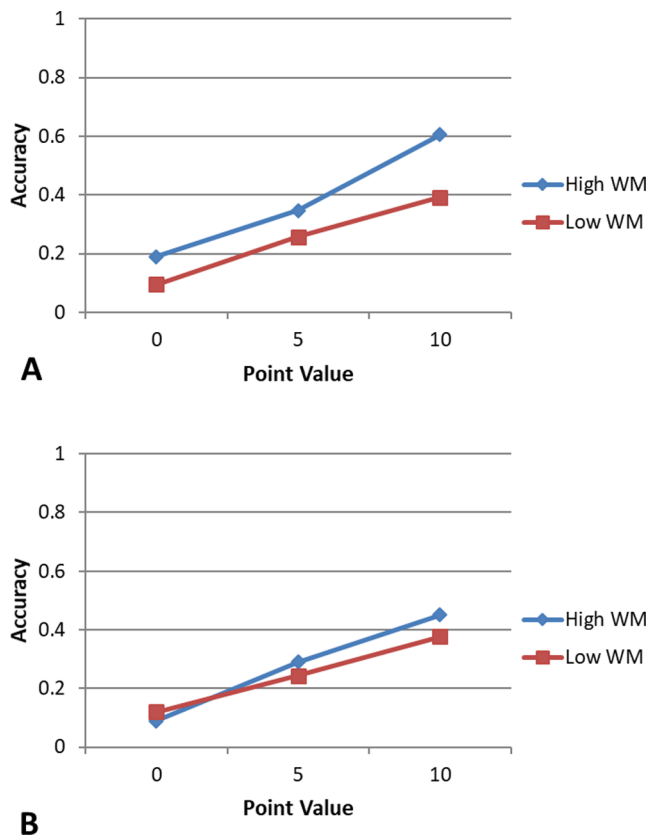

Fig. 4. Accuracy at each point value for those with high and low working memory in Experiments 2a and 2b.

Table 2
Correlations between working memory subtests combined across all experiments.

	Operation	Symmetry
Operation	–	
Symmetry	0.48	–
Reading	0.59	0.40

with selectivity, (r = .181, t(387) = 3.619, p < .001, BF_{10} = 33.989), with higher working memory predicting greater prioritization. Fig. 6 shows these relationships collapsed across all experiments. We also considered the possibility that selectivity varied in part by trading off with recall accuracy. Fig. 7 shows the correlation between selectivity and recall accuracy, and shows that, across subjects, higher recall did not reliably predict worse selectivity, r = −.049. Fig. 8 provides another way to visualize prioritization, by comparing recall accuracy to working memory at each point value. If high working memory subjects are more selective, we should see this effect reflected in the slope; a steeper slope for high value word pairs would indicate that the boost in recall is concentrated in the higher-value items. If selectivity does not vary by WM, we should see three parallel slopes. Consistent with the selectivity analysis, a regression that allowed for different slopes/intercepts at each point-value indicated steeper slopes for high value word-pairs compared to low (0) point word-pairs (t(1194) = 4.463, p < .01).

General Discussion

Across four experiments, we have reaffirmed two well-known findings: high working memory predicts better long-term memory in experiments, working memory correlated with recall (r = .285, t(398) = 5.926, p < .001, $BF_{10} = 1.015 \times 10^6$), with higher working memory predicting more accuracy. Working memory also correlated recall, and people can prioritize study to selectively remember valuable materials. We also asked a novel question—can working memory predict this ability to prioritize? Based on an analysis of all of the data included here, the answer appears to be yes—though the size of the effect overall is very small, and it appeared only unreliably across experiments.

Part of the reason for this effect size may simply be due to lower reliability of selectivity as a measure. Reliability acts as a ceiling on the

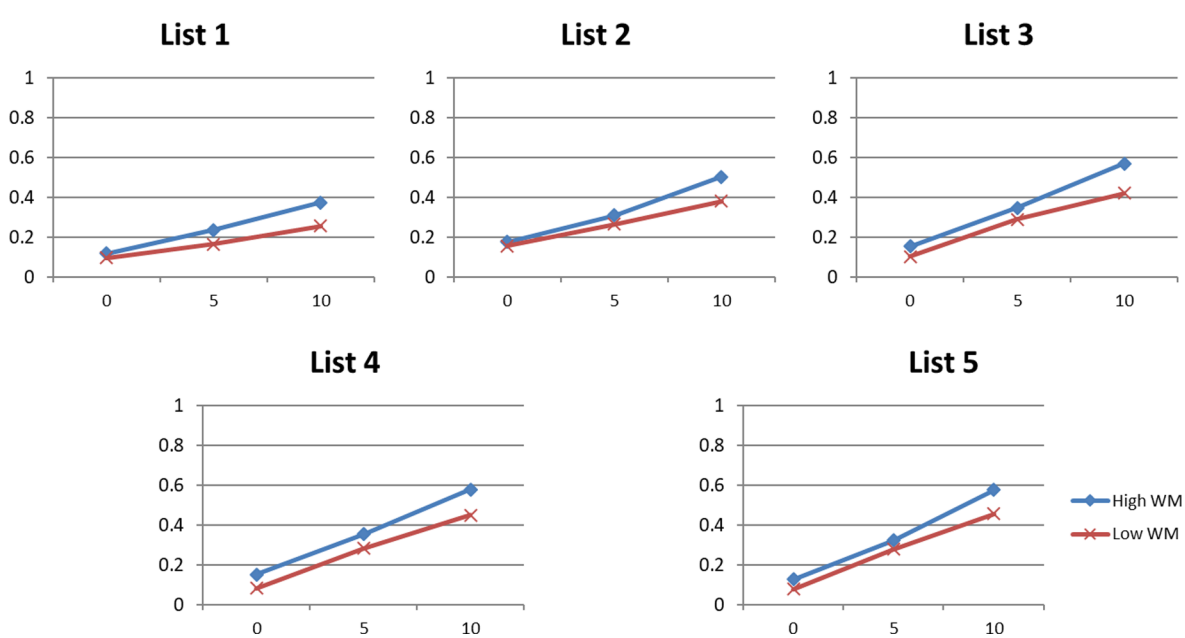

Fig. 5. Accuracy for 0, 5, and 10 point word pairs for each list in Experiments 2a and 2b.

(*footnote continued*)
correlations were within ~.05 of each other and the omnibus correlation matrix.

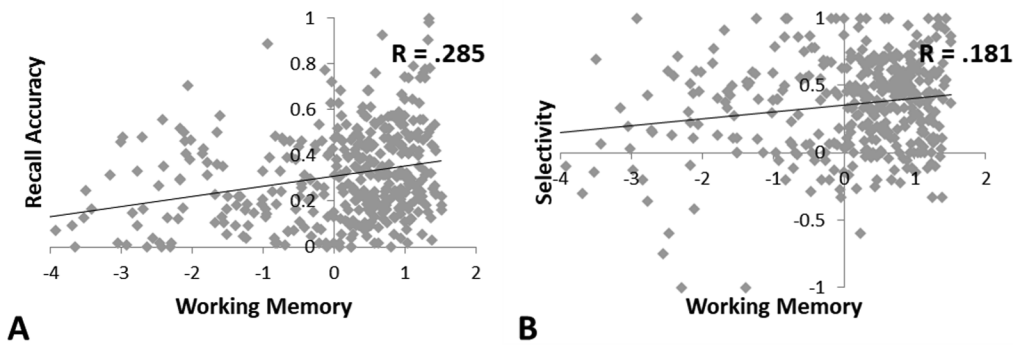

Fig. 6. The relationship between working memory and cued recall accuracy (A) and selectivity (B) across all experiments.

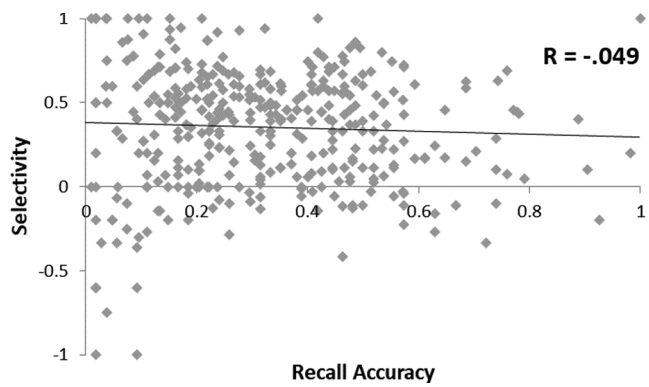

Fig. 7. The relationship between selectivity and recall across all experiments.

magnitude of a correlation. If reliability is low, even a moderately strong effect might appear inconsistent. On the other hand, even when reliability was a quite high 0.8 in Exp. 2a, the relationship between WM and selectivity remained modest. This small effect is somewhat surprising, given the ample evidence for superior strategy use in long-term retrieval from semantic as well as autobiographical memory for those with high WM (Unsworth et al., 2013; Unsworth, Spillers, & Brewer, 2012).

One option for subjects in our experiments is simply to ignore 0 point items, since the recall of those items can never improve their total score. There is evidence that working memory can aid in the suppression of such irrelevant material, including the fact that HWM predicts more substantial directed forgetting (Delaney & Sahakyan, 2007). Working memory also predicts suppression in broader cognitive tasks, including anti-saccade performance (Kane, Bleckley, Conway, & Engle, 2001), and dichotic listening (Conway, Cowan, & Bunting, 2001). In short, if the best performance in our task were purely a matter of keeping attention away from the 0 point items, then the above results all suggest HWM learners should be more capable of doing so. Yet the advantages were only modest in our task.

Another possibility is that high working memory confers benefits in the diversity of strategies possessed and ease with which they can be deployed, but not in the effectiveness with which they can be locally modulated over materials. Some strategies, like chaining, require each to-be-remembered word to be integrated into a continuing story; putting the story on hold mid-rehearsal in order to selectively ignore some items may be more difficult than weaving in those irrelevant pieces. A similar effect may be at work in item-method directed forgetting: Fawcett and Taylor (2008, 2012) showed that subjects are slower to respond to an unrelated probe following a forget cue than a remember cue, an effect that is consistent with the idea that directed forgetting is an *active*, not passive, process. If clearing the contents of working memory takes effort and the benefits of it are not large and obvious—it is unlikely that subjects have rich theories about reducing intra-list interference—then subjects may not always make the effort to avoid encoding low-priority material.

A similar unwillingness to shift strategies on an item-by-item basis is evident in recognition, where people are reluctant to employ different criteria for different classes of items that are intermixed on a test list. Even when one category of materials is more well learned than another—a manipulation that normally leads to an elevated recognition criterion and lower false alarms—subjects do not employ that higher criterion when the classes of items are intermixed (Stretch & Wixted, 1998). Only when the item itself guides subjects straightforwardly to the class to which it belongs are item-by-item changes in criterion evident (Benjamin, 2001; Benjamin & Bawa, 2004; Rotello & Macmillan, 2008; Verde & Rotello, 2007).

It is also worth reconsidering the strategy-affordance view of Bailey et al. (2008). Because prioritization is not a strategy that can be shared across the working memory tasks and value-directed remembering, it is less likely to mediate the known relationship between working memory and recall. Even if strategy use is responsible for some of the

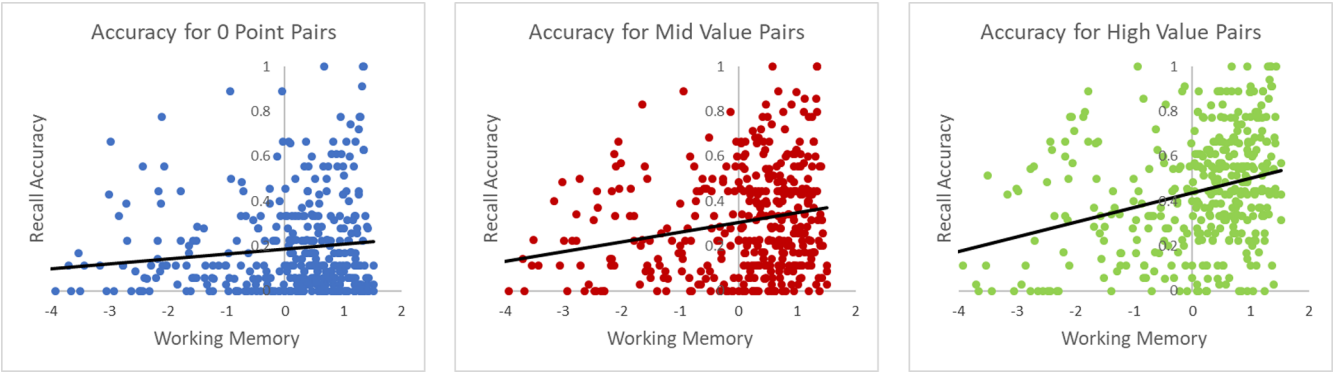

Fig. 8. The relationship between working memory and cued recall accuracy for words at each point value, across all experiments.

relationship between recall and working memory, it may not reveal itself in a task-unique strategy like the ability to prioritize.

We return here to a recent paper on strategy use in value-directed remembering (Robison & Unsworth, 2017), which directly addressed the relationship between WM and free recall by *giving* participants an effective strategy to use in a free recall variant of the task, and seeing if WM differences disappeared. In their task, participants received lists of *only* point values, and had to click on a point value to access the word tied to it. After two minutes had elapsed they took a free recall test. Participants here had exceptional control during study—over both study time and the actual appearance of words, and the *best* performance arose when participants ignored a subset of the lowest value words completely. The second experiment included instructions that directly provided this strategy to participants. Under those instructions, differences in the VDR task vanished.

On the basis of this finding, the authors argued that although low WM participants may spontaneously use worse strategies, those ineffective strategies alone are insufficient as a mediating variable. Our own modest effects of working memory on selectivity square with this idea—in none of our experiments are the effects large enough to mediate any sizable proportion of the relationship between working memory and recall. A recent paper is yet more pessimistic about the relationship between WM and prioritization: within delayed free recall, there was no reliable correlation ($r = -.056$) between selectivity and WM (Miller, Gross, & Unsworth, 2019). Instead, both high and low WM participants appeared to effectively prioritize items for study.

One important difference between the previous papers and our current task is that our study specifically required participants to exert control only at *encoding*, and not at retrieval. During retrieval, items are in competition: the need to report items serially is a bottleneck that forces prioritization, as memory can decay and interference can accrue. In fact, Castel, Murayama, Friedman, McGillivray, and Link (2013), using the same method as Robison and Unsworth, found that some older adults' achieved a gain in selectivity by controlling output order—by studying high value words late and outputting them first, they engineered a recency dump that minimized interference from those items. Moreover, because rememberers control the pace of retrieval, they can take additional time to exercise control or use a more involved (retrieval) strategy. What is surprising is that, in both our study and that of Robison and Unsworth (2017), there were small effects of working memory on selectivity, and similar WM disparities when participants were left to their own devices. Even with a considerable degree of control over their study, high-WM participants did not demonstrate a substantial advantage in prioritization. Yet we do see in our study quite clear evidence that, across all levels of WM, participants did get significantly better at prioritizing over the course of the task.

This latter result in particular is suggestive, since it shows that prioritization is not bound by a hard limit in cognitive ability or capacity but rather is subject to rapid and substantial improvement with feedback and experience. That both groups can improve over time (and that there is only a modest and inconsistent correlation between WM and prioritization) suggests that the selective encoding necessary for prioritization is a skill that does not tax a limited working memory, but rather reflects a lack of experience.

Additional research on prioritization lends credence to the idea that prioritization may not depend on WM resources. Middlebrooks, Kerr and Castel (2017) conducted a value-directed remembering task while manipulating whether attention was divided during encoding. There are many conditions in which the imposition of a cognitive load at encoding removes the advantage in recall for high-WM participants (Engle, Cantor, & Carullo, 1992); if prioritizing items relies on this same resource, we would also expect prioritization to suffer under such a load. Instead, prioritization was unaffected by the division of attention. In fact, although overall recall dropped, memory for the most valuable items stayed constant across conditions. It would not have been surprising if dividing attention undercut prioritization. Other aspects of strategic encoding, such as deeper, more semantic processing (that high-WM subjects more commonly self-report using) are particularly disrupted by divided attention. Yet prioritization appears robust to it, as well as present in low WM subjects here.

That *both* groups improved with experience in prioritizing, despite only modest differences in WM, is a heartening result. This result suggests that many subjects learned not only how to remember words better, but also how to be more strategic in their study choices and focus on the items they deemed most important. Effective strategies undoubtedly play a role in supporting good and incisive use of memory, and they may be partly related to working memory capacity, as shown here, but those strategies clearly do not account for much of the high and consistently shown relationship between working memory and recall.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2019.104032.

References

Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex cognition? Testing the strategy affordance hypothesis. *Memory & Cognition, 36*(8), 1383–1390.

Benjamin, A. S., & Bjork, R. A. (2000). On the relationship between recognition speed and accuracy for words rehearsed via rote versus elaborative rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 638–648.

Benjamin, A. S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 941–947.

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31*, 297–305.

Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory & Language, 51*, 159–172.

Castel, A., Benjamin, A. S., Craik, F. I., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition, 30*(7), 1078–1085.

Castel, A. D., Balota, D. A., & McCabe, D. P. (2009). Memory efficiency and the strategic control of attention at encoding: Impairments of value-directed remembering in Alzheimer's disease. *Neuropsychology, 23*(3), 297–306.

Castel, A. D., Humphreys, K. L., Lee, S. S., Galván, A., Balota, D. A., & McCabe, D. P. (2011). The development of memory efficiency and value-directed remembering across the lifespan: A cross-sectional study of memory and selectivity. *Developmental Psychology, 47*, 1553–1564.

Castel, A. D., Murayama, K., Friedman, M. C., McGillivray, S., & Link, I. (2013). Selecting valuable information to remember: Age-related differences and similarities in self-regulated learning. *Psychology and Aging, 28*(1), 232–242.

Clarkson-Smith, L., & Hartley, A. A. (1990). The game of bridge as an exercise in working memory and reasoning. *Journal of Gerontology, 45*, 233–238.

Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic bulletin & review, 8*(2), 331–335.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.

Delaney, P. F., & Sahakyan, L. (2007). Unexpected costs of high working memory capacity following directed forgetting and context change manipulations. *Memory & Cognition, 35*, 1074–1082.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: a test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 972–992.

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*, 309–331.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis.* Cambridge, MA: MIT press.

Finley, J. R., & Benjamin, A. S. (2012). Adaptive and qualitative changes in encoding strategy with experience: Evidence from the test-expectancy paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 632–652.

Fawcett, J. M., & Taylor, T. L. (2008). Forgetting is effortful: Evidence from reaction time probes in an item-method directed forgetting task. *Memory & Cognition, 36*(6), 1168–1181.

Fawcett, J. M., & Taylor, T. L. (2012). The control of working memory resources in intentional forgetting: Evidence from incidental probe word recognition. *Acta Psychologica, 139*(1), 84–90.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*(2), 169–183.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press, Clarendon Press.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition,*

32(3), 609–622.
McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition, 29*(1), 10–17.
Middlebrooks, C. D., Kerr, T., & Castel, A. D. (2017). Selectively distracted: Divided attention and memory for important information. *Psychological Science, 28*(8), 1103–1115.
Miller, A. L., Gross, M. P., & Unsworth, N. (2019). Individual differences in working memory capacity and long-term memory: The influence of intensity of attention to items at encoding as measured by pupil dilation. *Journal of Memory and Language, 104*, 25–42.
Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General, 145*(2), 200–219.
Nelson, D. L., McEvoy, C. L. & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/freeassociation/.
Robison, M. K., & Unsworth, N. (2017). Working memory capacity, strategic allocation of study time, and value-directed remembering. *Journal of Memory and Language, 93*, 231–244.
Rotello, C. M., & Macmillan, N. A. (2008). Response bias in recognition memory. In A. S. Benjamin, & B. H. Ross (Vol. Eds.), *The Psychology of Learning and Motivation: Skill and Strategy in Memory Use: Vol. 48*, (pp. 61–94). London: Academic Press.
Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237.
Sahakyan, L., Delaney, P. F., & Kelley, C. M. (2004). Self-evaluation as a moderating factor of strategy change in directed forgetting benefits. *Psychonomic Bulletin & Review, 11*(1), 131–136.
Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology-Learning Memory and Cognition, 26*(1), 204–221.
Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1379–1396.
Tullis, J. G., & Benjamin, A. S. (2012). The effectiveness of updating metacognitive knowledge in the elderly: Evidence from metamnemonic judgments of word frequency. *Psychology and Aging, 27*, 683–690.
Tullis, J. G., Fiechter, J. L., & Benjamin, A. S. (2019). The efficacy of learners' testing choices. *Journal of Experimental Psychology: Learning, Memory, and Cognition* (in press).
Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition, 41*, 429–442.
Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352–373.
Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory & Language, 49*, 446–468.
Unsworth, N. (2007). Individual differences in working memory capacity and episodic retrieval: Examining the dynamics of delayed and continuous distractor free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 1020–1034.
Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica, 134*, 16–28.
Unsworth, N. (2016). Working memory capacity and recall from long-term memory: Examining the influences of encoding strategies, study time allocation, search efficiency, and monitoring abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 50–61.
Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498–505.
Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17*, 635–654.
Unsworth, N., Spillers, G. J., & Brewer, G. A. (2012). The role of working memory capacity in autobiographical retrieval. Individual differences in strategic search. *Memory, 20*, 167–176.
Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: The role of controlled search. *Memory & Cognition, 41*, 241–254.
Verde, M. F., & Rotello, C. M. (2007). Memory strength and the decision process in recognition memory. *Memory & Cognition, 35*, 254–262.
Watkins, M. J., & Bloom, L. C. (1999). Selectivity in memory: An exploration of willful control over the remembering process. (Unpublished manuscript).