Article

Perceptual and Motor Skills
2022, Vol. 129(3) 378–398

DOI: 10.1177/00315125221076440
journals.sagepub.com/home/pms

Managing Increased Cognitive Load in a Guided Search

Josée Turcotte[1] and **Bruce Oddson**[2]

Abstract

In the Sternberg item recognition task and its variants, an individual's mean reaction time increases with the number of items to be retained in the memory set. An increase in reaction time has also been seen when a secondary task was added. The usual interpretation for this increased reaction time is that adding cognitive load makes tasks more difficult. In a series of three experiments, we manipulated cognitive load through increases in the memory set or through a second task. In each experiment, high cognitive load was associated with higher mean response times but a reduced slope, based on the target position in a series of probes. Thus, in a Sternberg task with multiple word targets and multiple word probes, participants searched more efficiently per probe under high load than under low load. This pattern was replicated with the addition of a working memory task requiring participants to calculate a cumulative price based on the price per target word item. By considering both initial response times and reaction time slopes in large memory sets, this study provides a challenge to the traditional interpretation of cognitive load effects on search performance.

Keywords

management, memory, working memory, cognitive load, Sternberg's task

---

[1] Faculty of Arts, Laurentian University, Sudbury, ON, Canada
[2] Faculty of Health and Education, Laurentian University, Sudbury, ON, Canada

Corresponding Author:
Josée Turcotte, Faculty of Arts, Laurentian University, 935 Ramsey Lake Road, Sudbury, ON P3E 2C6, Canada.
Email: jturcotte@laurentian.ca

Introduction

In the Sternberg item recognition task and its variants, participants are first presented with a list of items to remember (memory set), then presented with a probe item, and then asked to indicate whether the probe was is in the initial memory set. In the classical Sternberg task, empirical research has clearly and consistently found that the size of the memory set matters. As the number of targets in the memory set increases, the participants' reaction times for each probe increase as well. With accuracy high and stable, a linear increase in the size of small memory sets has been associated with a linear increase in reaction times (see Sternberg, 2016, for a recent review of this research).

With larger memory sets the pattern of relationships between memory set and participants' reaction times has been less clear. Burrows and Okada (1975); Okada and Burrows (1978) found that, per increment in memory set size, the increase in reaction time per item was less for larger memory sets with 8–20 targets than for smaller sets with 2–6 targets. Surprisingly, this larger memory set phenomenon has never been well explained. As these authors first suggested, items in larger sets might be organized or grouped together differently than in smaller sets, so as to make memory processes for larger sets more efficient (for an opposing view, see Sternberg, 2016).

Why has such a well-studied memory paradigm not led to a clear understanding? First and foremost, a simple correlation between memory set size and reaction time does not constrain theories as complex as those that have been proposed. Second, reaction time is presumed to be a combination of different stages of mental processing (reading a probe, scanning memory, and decision-making), meaning that reaction time differences can only be loosely associated with memory models, and other dependent measures are needed as well. Underlying various models of memory search are questions about how memory sets might be represented. It is well established that the nature and complexity of memory stimuli matter (Sternberg, 2016) and that memory decision-making may be subject to both within and between participant variations in memory capacity and strategies (e.g., Corbin & Marquer, 2009, 2013).

Again, in Sternberg's paradigm tasks, the reaction time reflects the time needed to make a decision about the probe. In this situation it has been natural for researchers to conclude that cognitive load delays decision-making time and may influence the error rate. This attribution is presumptive, and there have been few investigations regarding which information processes are most affected. In principle, however, cognitive load may also include processes related to the perception of the probe; and, possibly, some part of reaction time reflects a re-processing and organizing of the memory set. This possibility has not been carefully examined. However, from an ecological perspective, memory processes are optimized. In everyday life, our use of memory is not to support a search through random lists of unrelated items, but to search lists that share some categorization and function. Thus, the way in which representations of memory sets are used and how memory is searched may interact in determining memory load and reaction time.

Some Search Parameters

Under the classic interpretation of memory scanning (Sternberg, 2016), each probe word (from a reading list) is compared to each target word (held in memory) in separate steps. The expected time to respond should increase with the size of the memory set multiplied by the number of word probes needed to locate a positive match between the memory set and the word probes. Both memory set size and the number of probes jointly determine the response time. If every memory set and word probe comparison is independent and sequential, then the response time will mainly reflect the effects of memory set size and the target position of the memory set in the probe list. But response times actually reflect other factors that may be constant, even as the number of memory comparisons increase. In negative trials, although the memory set size and the number of probe words are the same as in positive trials, the number of comparisons to be made is maximal. Note, however, that the mean time per item for a correct rejection is not necessarily related to the mean time per item for a correct positive decision, as a rejection decision may use a separate mental process (for example, see Johns & Mewhort, 2002).

One of the classic investigations of memory set size in Sternberg tasks was conducted by Burrows and Okada (1975); Okada and Burrows (1978). As noted above, in their data (see also Brigg, 1974) the marginal impact of increases in memory set size was reduced for larger memory set sizes. This is surprising since a large number of targets to remember would be expected to impose constraints on the search process, relative to tasks with smaller memory loads. With smaller (sub-span) memory loads there is a presumption that working memory insures a fast and complete comparison process with each probe item. When memory load is above the working memory span, the presumption has been that the memory comparison process becomes more difficult and time consuming. All other things being equal, difficulty through increased cognitive load would be expected to lead to slower reaction times.

In the broad memory literature, the classic interpretation has been that supra-span memory loads are not constrained by working memory limits, and our representations are flexible enough to meet the challenge. The way this is handled has been described differently, depending on the investigator’s underlying theory of memory. For example, global matching theories of memory have held that the memory set for targets is stored in a composite or compound form, such that, for each probe, one comparison is made with the memory set, and large memory sets make the comparison less efficient (Murdock, 1982, 1983, 1995). On the other hand, local matching memory theories (e.g., Shiffrin & Raaijmakers, 1992; Shiffrin & Steyvers, 1997) have suggested that all items are individually represented in memory, meaning that comparisons are made in parallel, and memory set size effect can be explained by inter-item interference. While both of these theories imply that the size of the memory set has an impact on decision-making time, both theories make flexible predictions, and it is difficult to put these theories to a definitive test.

The distinction between sub-span and supra-span tasks also relates to theories of working memory. In principle, working memory capacity is very limited. The memory set of targets, the process of reading the probe list, and the process of making decisions about the match between the probe and one of the targets in memory all share demands on this limited working memory capacity. For sub-span tasks the performance for each probe to make a target memory comparison can be expected to be faster than for supra-span performance. Supra-span conditions exceed working memory capacity, and no memory theory suggests that comparisons between targets and probes would become faster. Rather, one would expect that a high memory load would lead to more errors or more difficulty managing the task. In point of fact, supra-span memory loads have been routinely used as cognitive loads in precisely this manner to investigate claims about working memory.

Finally, consider the process of reading the list of probe words. The deployment of attention over lists of text is highly efficient in fluent readers. Probes may be read for comparison individually and sequentially, or in small groups. When probes are read individually or sequentially, whatever variability reading time contributes to this process should be additive and further increase the time to find a target on the list. When probes are read in small groups, the situation is potentially more complex. Logically, the contribution of reading time should increase relatively slowly as the number of probes to consider increases, though this process places an additional load on the memory system, which might be reflected in decision-making time as well. Thus, it seems that reading processes should probably be considered along with other task dimensions.

This brings us to the principal question of the present study. Given the need for memory of both a list of targets and of a list of probes, one of which may contain a target, how do participants manage their search process? In particular, do the sizes of the memory set and the probe list interact in their impact on participants' response times? It should be possible, in empirical research, to design ways to constrain variables associated with different theories of how memory is managed for ecologically valid (i.e., real world) tasks. In this study, we distinguished between changes in mean reaction times for participants' initial responses (the intercept, identified in the following experiments with a main effect of memory set size) and changes in the slope of the participants' reaction times over the number of probes read before a match could be found (the slope, identified though an interaction between memory set size and the location of the target in the list of probes).

Experiment 1

Maintaining a gluten-free diet is a real-world memory challenge. To avoid exposure from processed foods, people must memorize lists of up to 26 gluten containing ingredients (wheat, etc.). They then need to check product ingredient lists to make sure none from that first list are present. This actual circumstance, faced by people with celiac disease, has not been studied in situ. Difficulties with this task have generally been attributed to questions of assiduity and motivation (e.g., Barberis, et al., 2019; Rubio-Tapia, et al., 2013). We simulated this challenge using a smaller number of targets, in hopes of recreating the dynamics of supra-span target searching with a list that can be feasibly memorized within a short period. Logically, cognitive load theory (Sweller, 1988) would predict that the greater the number of target words to memorize (increased cognitive load), the lower participants' recognition accuracy and the longer the task would take. Also, accuracy would likely be higher and reaction time would likely be shorter for positive trials when the target was present (and might be found sooner in the mental memory set) in comparison to negative trials when the target was absent (and the entire memory set had to be searched to make this determination).

Experiment 1 Method

Research Design and Approval

In this experiment, we used a list of food ingredients in a product as our stimuli and varied the memory trial type (positive and negative) and the number of food ingredients. We varied the number of ingredients (probes) within participants, and the number of targets to search for between participants. We also manipulated the provision of product names between participants; but, after further analyses showed that the presence or absence of product names generated the same pattern of results, we collapsed these conditions. All participants signed a written informed consent form before participating in this research, and this study received research ethics approval from the university's review board.

Participants

Fifty-four (39 women and 15 men) Laurentian University students between the ages of 18 and 39 (*M* = 22.17, *SD* = 5.52) volunteered to participate in this experiment. Eligible students (for example, those taking Introduction to Psychology) received a course bonus point for their participation. Their number of years of education varied from 12–22 (*M* = 14.62, *SD* = 1.75). A post-experimental questionnaire revealed that 13 participants had heard about celiac disease and that one (participant #11) had celiac disease; other participants had knowledge of celiac disease through contact with someone living with it.

Stimuli and Materials

We presented target words and word stimuli using E-prime (Psychology Software Tools, Inc, Philadelphia, USA). Eight target ingredients containing gluten had been previously selected: wheat, barley, rye, durum, spelt, malt, and kamut, along with eight gluten-free non-targets for the control condition: quinoa, millet, tapioca, potato flour, arrowroot, buckwheat, bean, and chickpea. A total of 384 stimuli (food product ingredient lists) were generated to allow full counterbalancing of stimuli between participants.

We presented a target word list of either two or eight words all at once to the participant. Next, we presented a list to be read of either short probe words (4–10 ingredients) or long probe words (11–25 ingredients). Still showing the probe list, we then asked participants if any of the target words were present in the actual probe list. Each participant responded to 96 trials, with 50% of trials considered positive trials in which a target word was present in the probe list and 50% of trials considered negative trials in which none of the target words were present in the probe list.

Procedure

Participants completed a pre-experimental questionnaire and then performed a practice memory test before engaging in the main experimental trials. In the practice session there was a single target, "walnut," that participants searched for when presented with six randomly ordered food product ingredient lists (probe list), half of which contained this target word and half of which did not contain it. All practice stimuli were different from stimuli used later in experimental trials. Participants selected their positive responses with their dominant hand, whereas they used their non-dominant hand to give negatives responses.

For the main experiment trials, participants saw either two target words for 10 seconds or eight target words for 40 seconds. We asked the participants to memorize those targets (the memory set). Next, we showed participants a list of probe words to be read, and we then asked if any of the target words were present in the probe list. This list disappeared once the participant had given an answer. Then, participants pressed the spacebar to go to the next probe list—thereby self-selecting the length of the break interval between lists. An additional break occurred between each block of 24 stimuli. After all four test blocks were completed, we tested the participants' memory for the initial target word list using forced choice recognition. Finally, we asked participants to complete a post-experiment questionnaire regarding their knowledge about diabetes and celiac disease, both of which are medical conditions treated in whole or in part with a prescribed medical diet.

Data Analyses

We conducted all analyses using STATA 16.1 (StataCorp, 2019). We analyzed participant accuracies using a mixed effects logistic regression model, nested by participant. We analyzed participant reaction times using generalized regression assuming a gamma distribution. This distribution naturally describes non-negative data with positive skew, as expected for reaction times. Statistical significance was set at $p < .05$.

For data analysis in this paper, we only include the following: (a) reaction time data for the trials in which the reading task was accurate and in which the target recognition task, after the reading task, was successful and (b) accuracy rate data for the trials in which the target was successfully recognized in the target recognition task. But equivalent results were found when this restriction was relaxed and even when inaccurate responses were included.

Experiment 1 Results

Recognition of Target Words. On our tests of the participants' forced choice recognition memory for the target words, following all experimental trials, the participants' recognition memory was excellent (99%) for the 2-item condition and good (71%) for the 8-item condition. Considering that some targets were not recognized, it was difficult for the participants to identify them correctly in the reading trials, and this may have affected their reaction time. We excluded those trials from our analysis of the reading task.

Reading Trials

The participants' accuracy on positive trials (i.e., those with target words present in the probe list) was higher when searching for two targets (88%) then when searching for eight targets (72%), $z = 4.1$, $p <.001$. Accuracy was also higher when the list to be read had few probe words (84%) as opposed to many (79%), $z = 2.38$, $p = .017$. The interaction between memory set size and probe list size on positive trials was not statistically significant, $p = .06$, and there was no significant linear effect of probe position in the list of probes, $p = .193$, nor any quadratic effect, $p =.227$.

Participants' accuracy on negative trials (i.e., those for which no target words were present) was near perfect for two target words (97%) and good for eight target words (77%), $z = 5.65$, $p < .001$. There was no significant accuracy difference between lists with smaller or larger numbers of probes, $z = 1.29$, $p = .197$; and, on negative trials, there was no significant interaction between target list size and probe list size, $p = .961$.

Regarding participants' reaction times on positive trials of the reading trials, we analyzed them using mixed effect modeling based on a gamma distribution (this was a better fit than a model assuming normally distributed errors, $p < .001$). Participants responded faster when the memory set or target word list size was two than when it was eight, $z = 3.6$, $p < .001$. They also reacted faster and showed a linear effect with a smaller versus a larger number of probes, $z = 8.7$, $p < .001$. There was a significant interaction effect between number of target words in the memory sets and the number of probe words to be read, $z = -3.3$, $p < .001$. The position of the target amongst the probes was very important, $z = 7.77$, $p < .001$ but there is also a quadratic effect of position, $\mathrm{z} = -3.1$, $p = .002$. On post-hoc testing, the size of the memory set interacted significantly with the position of the target amongst the probes, $\mathrm{z} = -2.8$, $p = .005$. Participants' responses were more affected by the number of probes to consider when the memory set size was small than when it was high (Figure 1).

For participants' reaction times on the negative trials of the reading task, we found no significant difference between trials with two versus eight target words, $z = .47$,

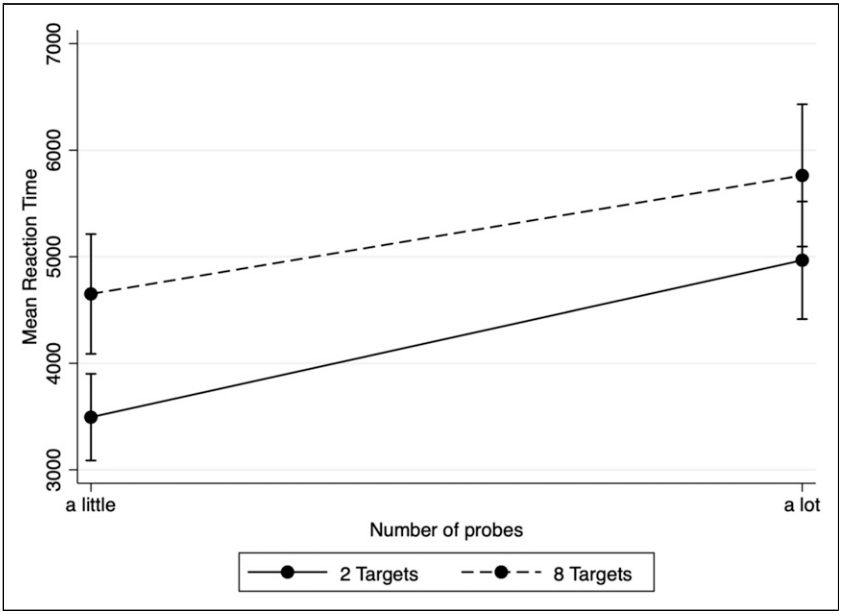

Figure 1. Memory Set Size with Number of Probes.

$p = .64$. Contrary to the accuracy-related findings, there was a significant probe size effect on reaction times, $z = 34.32$, $p < .001$, such that response times were faster when there were fewer probes to read. There was no significant interaction effect between number of probes and number of target words (memory set size), $z = -1.61$, $p = .107$.

Experiment 1 Discussion

The search process in this study involved (a) maintaining in working memory a memory set of target words, (b) reading and attending to a list of presented word probes to be compared with recalled targets, and (c) making an appropriate decision. Participants may have engaged in a series of self-terminating searches of the probes or some parallel comparison of presented word probes with recalled targets. Each of these strategies was expected to affect the participants' response rates. In Experiment 1, we varied the size of the memory set of target words (i.e., two or eight) and the number of presented word probes (i.e., targets and non-targets) in the reading task. To help assure that participants retained the memory set, we used a recognition task.

We first found that participants almost always perfectly retained two memory targets, but they were more challenged by eight targets. Our decision to ignore trials on which targets were not recognized may have been overly strict, since at least some of these trials led to correct positive responses. But we preferred to narrow our investigation to search processes over remembered targets rather than all presented targets.

Both the size of the memory set and the number of probes in the reading task affected participants' accuracy (even on negative trials) whereas the position of a target amongst the presented probes had no effect on response accuracy.

On positive trials, the participants' rate of responding was much faster for two (vs. eight) memory targets. However, size of the memory set significantly interacted with the number of probes in the reading task, such that the number of probes made less difference in participants' reaction times when the memory set was large and more difference in reaction time when the memory set was small. For positive trials, the number of probes and the position of the targets were not independent; both significantly contributed to reaction times for accurate responses.

Negative trials meant that no target from the memory set matched any of the probe ingredients in the reading task. Thus, these trials had longer reaction times than positive trials, because there was no possibility of early detection of a target that would self-terminate the search process. In this context, our finding of no memory set size effect on reaction time posed more questions than it resolved. If negative decisions were analogous to positive decisions, we would expect to see a strong memory set size effect. Instead, we have two main interpretations of these data. First, when working with eight targets in the memory set, participants accepted a higher probability of error and worked just as quickly as with small sets. Second, with eight targets, participants tended to rescan probe reading lists until they were comfortable that they had omitted no targets. As this experiment made no attempt to guide or study participants' strategies, both of these (and other) possibilities are feasible explanations for our findings.

The main surprise in Experiment 1 was that memory set size and number of probes in the reading task had a sub-additive effect on reaction time. That is, small (two target) memory sets led to slower response rates than large (eight target) sets. We rechecked and repeated this finding using several models, including a generalized gamma survival model and several competing risks regression models. As described in the introduction to this paper, if participants read the probe list and engaged in a repeated serial search of the memory set, the reaction times should have been four times higher with eight than with only two targets in the memory set. Instead, we found a reduced slope (reaction time per probe word) with eight targets. Similarly, if the participants had used sequential independent comparisons, their reaction times would be expected to have increased in proportion to the number of probe words they read. We found increased reaction time variance, but the improvement in fit resulting from use of a gamma model for reaction time as compared with a normal (or heteroscedastic mixed) model suggests that the change in variance reflected a change in the shape (including skewness) of the data.

An implication of these findings is that participants considered items in the target memory set in one or more groups, rather than separately. While this possibility fits with some memory theories (for example global matching models like TODAM; Murdock, 1982), it was certainly not predicted. The efficiency we observed with eight target words in reaction time per probe word (slope), coupled with participants' delayed starting times, suggests that, with larger memory sets, participants may have engaged in additional processing before starting. Unfortunately, our experimental design did not inform us regarding the nature of information processing that led to the participants' improved efficiency with the larger memory set. Accordingly, we took up these issues in subsequent experiments.

Experiment 1 findings might be interpreted, in part, as a speed versus accuracy trade-off in which participants with eight target items concerned themselves less with accuracy for some items to gain speed. However, the reduced reaction time slope we found with eight items in the memory set mitigates this as the only explanation for our findings. These slopes would be consistent with participants having considered only one or two targets, or searching for only a few targets in this eight target memory set condition would have led to much lower accuracy than we observed. Second, guessing or risky decision-making strategies should have led to a poorer performance than we observed on negative trials with the long probe lists.

Among the relevant design limitations of Experiment 1 are that the size of the memory set varied between, but not within, participants. Second, our means of presenting the probe words introduced the possibility that participants did not read them serially. Finally, our use of within-span and supra-span memory set conditions was not ideal for testing the interaction between memory set size and probe list length because there was no guarantee that participants used the same memory processes in both conditions.

Experiment 2

We designed Experiment 2 to examine the interaction between the effects of memory set and probe list size when the memory sets were supra-span. We compared lists of eight target words with lists of 12 and 16 target words. Since a weakness in Experiment 1 was relatively lower participants' accuracy with eight targets, Experiment 2 investigated the challenge of supra-span lists while altering the practice protocol to stabilize participants' accuracy at a high level and reduce the possibility that participants would make major shifts in criteria for memory decisions. Each participant had to first recognize all the words in the memory set during a recognition task before completing the reading task. As for Experiment 1, all participants in Experiment 2 gave their written informed consent prior to participation, and a local institutional review board approved the research protocol.

Experiment 2 Method

Participants

Sixty-three (55 women and seven men) Laurentian University students between the ages of 17 and 38 (*M* = 20.77, *SD* = 4.82) volunteered to participate in Experiment 2. For one participant, the demographic information was lost. Participants received $10 or, if eligible (for example, those enrolled in Introduction to Psychology), course credit for their participation. Their number of years of the participants' education varied from 12–19 (*M* = 13.16, *SD* = 1.30). In the post-experimental questionnaire, 23 participants had heard about celiac disease or had some knowledge of gluten-free diets.

Stimuli

In general, we used the same material in Experiment 2 as in 1, except that in Experiment 2, we doubled the number of trials in each block from 24 to 48, yielding a total of 192 stimuli in each block per participant. This change also required the creation of more food product ingredient lists and the creation of conditions with 16 target words. To preserve counterbalancing as much as possible, we used 24 target ingredients (wheat, atta, bulgur, couscous, pasta, triticale, malt, kamut, durum, semolina, graham, farina, barley, rye, oat, spelt, seitan, seasoning einkorn, dinkel, beer, ale, farro, and emmer) and 24 gluten-free non-targets (for the control condition). For each participant, half of the probe lists were long (15–21 ingredients) and half were short (4–10 ingredients). For all of participants, we provided product names at the top of the probe lists.

Procedure

The procedure was identical to that of Experiment 1 except that, before beginning the reading task, participants completed a learning phase of the to-be-memorized target words, starting with the presentation of the to-be-memorized targets list (5 seconds per target) followed by completing a forced recognition task. If the target words were not perfectly identified we repeated this learning phase until the participant recognized all target words.

Data Analyses

Our methods of data analyses were the same in Experiment 2 as in Experiment 1.

Experiment 2 Results

Recognition of Target Words

Participants' accuracy for forced choice recognition of the memory set (target words) following all experimental trials was high for all probe sets: 8-words (96%), 12-words (90%), and 16-words (91%). When a target word was not retained by a participant, we omitted analysis of the participant's reaction times for the trials using the corresponding target in the reading task.

Reading Task

On positive trials of the reading task, mean accuracy of recalled target words was 89%. There was no significant difference in participants' accuracy between trials with 8 versus 12 targets, $z = .29$, $p = .78$ or trials with 8 versus 16 targets, $z = 1.51$, $p = .13$. There was a significant difference in accuracy as a function of the number of probes to be read, such that accuracy decreased on trials with long versus short lists of probe words, $z = 4.57$, $p < .001$.

Regarding participants' mean reaction times on positive trials of reading, there were no differences between conditions with 8 versus 12 target words, $z = 1.34$, $p = .18$, or 8 versus 16 target words, $z = 1.37$, $p = .17$. There was a significant main effect for varying the number of probe words, $z = 4.8$, $p < .001$, such that, as the number of probes increased, so did the participants' reaction times. The position of a target amongst the probes was important, both linearly (mean reaction times slowed with more probes to read, $z = 21.91$, $p < .001$) and quadratically, $z = -11.21$, $p < .001$ (this increase was less severe for very long lists than would be predicted by the linear trend).

There was a significant interaction effect between memory set size and number of probes, $p < .05$; reactions times showed a higher slope—increment in reaction time per target position—for eight target words than for 12 target words, $z = -3.1$, $p = .002$, and for 8 target words than for or 16 target words, $z = -2.87$, $p = .004$. These results are summarized in Figure 2.

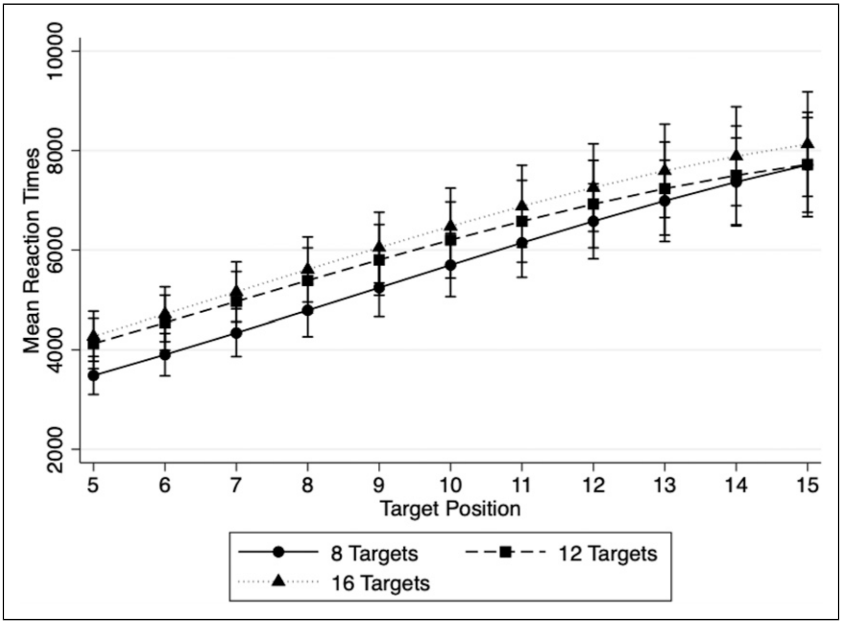

Figure 2. Reaction Times by Memory Set Size and Target Position.

Regarding participants' reaction times on negative trials of the reading task, there was no significant effect for number of targets; reaction times were not different for 8 versus 12 target words, $z = .73$, $p = .467$, or for 8 versus 16 target words, $z = 1.49$, $p = .15$. There was a significant effect of varying the number of probe words, $z = 52.54$, $p < .001$, but there was no significant interaction effect between the number of probe words for memory set sizes of 8 versus 12 target words, or 8 versus 16, respectively, $z = .77$, $p = .44$ and $z = -.02$, $p = .98$.

Experiment 2 Discussion

In Experiment 2, in which all memory sets were larger than the participants' working memory span, there was minimal impact of memory set size on participants' memory accuracy and reading task response times. Although the task was difficult, participants were able to do it. On trials in which our subsequent tests showed that the target word was recallable, neither positive nor negative trials showed an effect from the number of target words in the memory sets. Response times reflected the number of probes considered. On positive trials this is determined by the probes' locations. On negative trials this is the total number of probes. Most importantly, while there was a slight interaction effect between memory set size (number of target words) and the number of probes to read, for these larger memory sets in Experiment 2, this relationship was negative. Unexpectedly, participants who had to recall 12 or 16 target words in these memory sets were more efficient per item (i.e., showed better response time slopes across all trials) than those who had to recall only eight target words.

How does this finding relate to the participants' search strategies? There was no evidence that increases in participants' response times reflected serial processing of each target word, in turn. The absence of a memory set size effect supports the likelihood that participants considered a composite of all target items. The dominance of the target location over the number of probes to be considered suggests that participants read each probe sequentially and did not suffer interference from the total number of probes, as had been seen in Experiment 1.

Recall that, in Experiment 1 the interaction effect between memory set size (number of target words) and number of probes could be attributed to a qualitative difference between sub-span and supra-span memory loads, or simply to the quantitative difference between having to recall two versus eight target words. This distinction is of interest, because the presence of a qualitative effect from memory set size suggests that memory was managed differently as a function of memory set size. This possibility was strongly supported in Experiment 2 for which there was no strong interaction effect between memory set size and the size of the probe list for the longer supra-span lists in Experiment 2. In Burrows and Okada (1975); Okada and Burrows (1978), the impact of memory set size plateaued across their experimental conditions, and these investigators suggested that this impact might increase with the log of memory set size (see also Brigg, 1974). In the present experiment, the absence of an interaction effect suggests a very flat relationship between memory set size and the number of probes presented.

This finding goes against our intuitive understanding of the task. Task difficulty should have increased with the number of comparisons made; instead, participants searched the memory set exhaustively and maintained high accuracy and efficiency, even with large memory sets in Experiment 2. It is possible that a memory limit was no longer a bottleneck in the task. In this case, the observed reaction times seem to have primarily reflected the time it took to read the probes. It appears that reading and memory decision times were overlapping rather than separable and sequential. If the perceptual system fed the memory system at a relatively constant rate, and per probe word and memory comparisons did not delay this process, then we should not have expected memory set size to have any real impact on observed reaction times. Our results are consistent with and extend previous work that investigated supra-span tasks in the Sternberg paradigm. We conclude that our participants managed the search task quite efficiently, since a two-fold increase in targets and an increase in the number of probe words to target comparisons did not lead to an explosive increase in reaction time or errors. In Experiment 3 we sought to test whether the participants' ability to accommodate more targets in the memory set and more comparisons between the probes and targets in the memory set was related to a difference in cognitive processes or a qualitative difference between sub-span and supra-span memory searching.

Experiment 3

One possibility for participants' apparent efficiency when the number of probes and targets was high is that they engaged in additional cognitive processing to organize memory before beginning their searches. Alternatively, the results from Experiments 1 and 2 reflect some qualitative difference in how people recall small numbers of items, possibly within their working memory span, versus a supra-span memory set. To test these possibilities, we returned, in Experiment 3, to comparisons of single item, four- and 12-item sets. To highlight the potential role of working memory, we added to the memory task a cumulative price calculation working memory task by asking participants to memorize and sum the list price of probe words presented sequentially. That is, after every few trials, we asked participants to select a cumulative price of the items from amongst four options. This additional working memory task mimicked the additional cognitive load experienced by people engaged in real-world shopping tasks, when they add and estimate total costs as they shop.

Experiment 3 Method

Research Design

In Experiment 3, we varied trial types (positive, negative), number of probes and number of target words (with memory set sizes of 1, 4, or 12) and we varied the presence or absence of the cumulative price working memory task described above. All participants signed informed consent as before, and our research protocol received IRB approval as in Experiments 1 and 2.

Participants

Sixty (48 women and 12 men) Laurentian University students between the ages of 18 and 33 (*M* = 20.08, *SD* = 2.55) volunteered to participate in Experiment 3. Eligible students (those taking Introduction to Psychology) received a course bonus point for their participation. Their number of years of education varied from 12–16 (*M* = 13.17, *SD* = 1.20). In the post-experimental questionnaire, 17 participants reported having heard about celiac disease, while eight knew (directly or indirectly) someone with celiac disease.

Stimuli and Procedure

Our general procedure and materials in Experiment 3 matched those in Experiment 2, with a few changes. As before, there was a learning phase for target words before the reading task, and the reading task was completed in two blocks of 48 trials.

However, in Experiment 3, there were 1, 4, or 12 target words to memorize, with words drawn from this word pool: durum, malt, barley, oat, einkorn, pasta, atta, farina, seitan, bulgur, farro and beer. These words were matched with 12 gluten-free non-target words. In each block, half of the probe word lists were long (15–21 ingredients) and half were short (4–10). Also, in Experiment 3, we either added or did not add a cumulative price working memory task to one block of trials. For each trial with this task, a price (ranging from \$2.00–\$4.50) appeared below the ingredients list (probes) of a product. Participants had to determine the presence or absence of target words in the probe list as before, but they also had to recall the product's price, and after every four trials, they had to select the correct cumulative price from among four options (ranging between \$11 and \$15.75).

Experiment 3 Results

Recognition of Target Words

Participants' forced choice recognition of target words within the memory set was 95% and did not differ over conditions with differing numbers of target words.

Reading Task

Accuracy Rates and Positive Trials. On positive trials, the participants' accuracy rate with a single target word in the memory set (87%) did not differ significantly from their rate with four target words in the memory set (84%), $z = -.53$, $p = .59$. Their accuracy rate with 12 target words in the memory set (76%) was lower than their rate with a single target word, $z = -2.82$, $p < .01$, but it was not significantly different from their rate with four target words. Accuracy was negatively associated with the number of probes to read, $z = -3.53$, $p < .01$, meaning that, as the number of probes increased, participants' accuracy rate diminished. There was no significant interaction effect on accuracy between number of probes and memory set size when comparing memory sets of 1 and 4, $z = .11$, $p = .91$, but there was a near significant interaction effect between number of probe words and memory set size with memory sets of 1 and 12 target words, $z = 1.95$, $p = .05$. There was no significant linear effect for target position, $z = 1.05$, $p = .29$, nor was there a significant quadratic effect, $z = 1.15$, $p = .25$.

For positive trials, there was a significant interaction effect on accuracy between the presence of the cumulative price working memory task and memory set sizes, such that responses to single targets were more accurate with the working memory task present (88% vs. 86%), $p < .01$. There was no significant interaction effect between the working memory task and memory set size for memory sets of four or 12 items, $p = .08$ and $p = .09$, respectively. Regarding target position and working memory task presence, there was a significant interaction effect between them, $p = .02$. Without the cumulative price working memory task, there was a slight increase in accuracy over target positions (from 82% at position five to 84% at position 15), with the opposite pattern when the working memory task was present (from 85% to 81%).

Accuracy Rates and Negative Trials. For negative trials, accuracy for single target words (95%) did not differ from that for four target words (97%), $z = .65$, $p = .52$, but there were statistically significant lower accuracy rates between 12 target words (84%) compared to a single target, $z = -2.19$, $p = .03$. The number of probes to read did not have a significant impact on accuracy, $z = -.6$, $p = .55$ and there was no interaction effect between number of probes and memory set size among negative trials (1 vs. 4, $z = .12$, $p = .91$; 1 vs. 12, $z = -.86$, $p = .39$).

Reaction Times and Positive Trials. Figure 3 presents the participants' mean reaction times as a function of the target word position amongst the probes and the number of target words in the memory set. As illustrated, there was a significant main effect of the cumulative price working memory task on reaction times for positive trials. Adding (vs. not adding) the working memory task was associated with significantly slower correct reaction times on positive trials (without working memory task, $M = 3975$ msec; with working memory task $M = 7752$ msec), $z = 40.04$, $p < .001$.

Regarding main effects of memory set size on reaction times for positive trials there was no reaction time difference between trials with a single target word versus four target words, $z = 1.46$, $p = .15$, but there was a significantly faster reaction time with a single target word versus 12 target words, $z = 6.07$, $p < .001$.

For reaction time slopes across positive trials, the slopes for single targets differed from 4 or 12, but there was a significant interaction between memory set size and cumulative price working memory task, $z = -9.14$, $p < .001$, such that slopes based on the target position in a series of probes with single targets decreased in the presence of

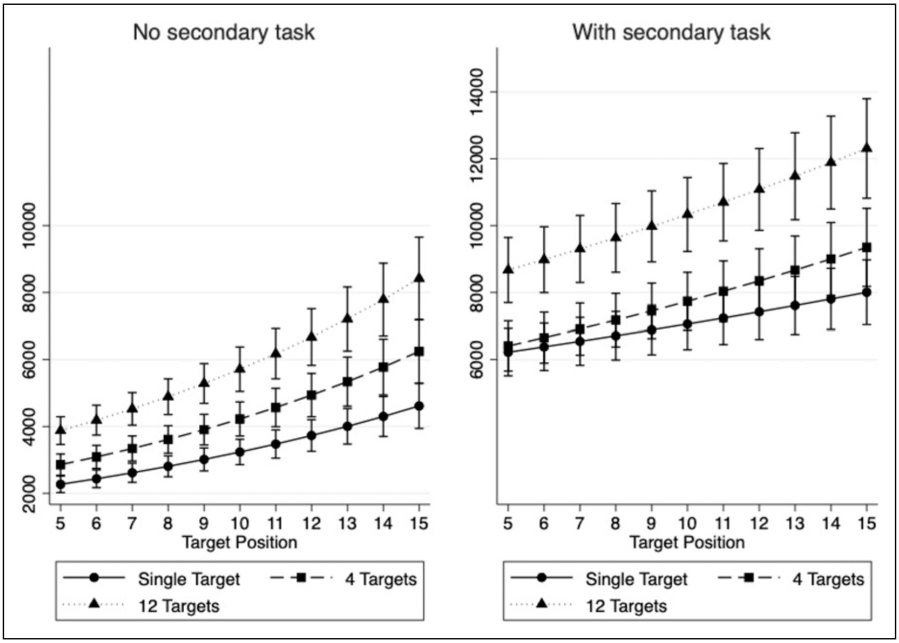

Figure 3. Reaction Times by Position and Working Memory Task.

the added cumulative price working memory task, as it did with four targets, $z = -12.6$, $p < .001$, and with 12 targets, $z = -13.62$, $p < .001$.

Reaction Time and Negative Trials. For negative trials the presence (vs. absence) of the cumulative price working memory task again led to a profound slowing effect on reaction time, $z = 37.16$, $p < .001$. The reaction times of negative trials in the single target word condition were not different than those with four target words in the memory set, $z = 1.16$, $p = .25$, but they were faster than negative trials with 12 target words, $z = 3.65$, $p < .001$. For reaction time slopes, there was a reduction in reaction time slopes when the working memory task was present (vs. absent) in memory sets consisting of a single target word, $z = -4.53$, $p < .001$, four target words, $z = -9.03$, $p < .001$, or 12 target words, $z = -8.69$, $p < .001$.

Experiment 3 Discussion

In Experiment 3, the main findings from Experiments 1 and 2 were replicated. The time to correctly respond did not increase with the number of comparisons to be made. Instead, there was an efficiency when many targets were to be searched for that was seen as a flatter slope over probe positions. In experiment 3 we added a working memory task and tested to see if it interacted with the slopes of a single, 4 or 12 targets.

In all cases the slopes were flattened, suggesting a change in the way search was managed.

Experiment 1 raised the possibility that differences in search efficiency might be related to working memory capacity; that searching for small numbers of targets was fundamentally different from searching for larger numbers of targets. However, in Experiment 3 both sub-span (1 and 4) and supra-span (12) target sets showed the same pattern when working memory load was manipulated.

General Discussion

In Experiment 1, we compared sub-span and supra-span memory sets. In Experiment 2, we compared three different supra-span memory sets. In Experiment 3, we manipulated both memory set size and the presence or absence of a cumulative price working memory task. Our critical findings were that, while participants more quickly identified target words when memory sets were small sets rather than large sets, memory set size effect interacted with the number of probes to be considered. This interaction effect was not, as might be expected, an additional speed or accuracy cost as the demands on memory and attention rose. Instead, we found a sub-additive interaction effect in which participants searching for target words from a large memory set showed less impact from having many probes words than those with a smaller memory set. Since the larger memory sets exceeded the participants' expected working memory span for a complex task, we concluded that participants found an efficient way to manage the increased perceptual, memory, and decision processing required by these tasks, instead of being overwhelmed by working memory demands.

In most previous work, investigators only studied reaction times to a single probe per trial. Accordingly, they could not evaluate the efficiency per probe participants have with multiple probes. In our results, at every point where we increased cognitive load (through increases in number of targets or the addition of a cumulative price working memory task), there was a mean increase in participants' reaction times but a reduction in their reaction time slope. While our research design cannot establish whether a qualitative difference in our participants' search processes with supra-span versus sub-span memory sets account for our findings, whatever process they used under higher cognitive loads, efficiency was preserved, even when additional demands were added.

The initial inspiration for our research was to simulate, with a real-world cognitive task, how people conduct memory searches in a day-to-day activity: searching for ingredients to avoid on labels (because of an allergy or a therapeutical diet to follow). We intuited that this task would become increasingly difficult as the number of target words in our memory sets increased. Surprisingly, we found that participants adapted and performed these ecologically relevant tasks with high, though imperfect, efficiency in their recall accuracy. We cannot discern whether their somewhat higher error rates with larger memory sets represent a trade-off for maintaining speed in perceiving and decision-making or are simply a function of having more opportunities to err.

Evidence presented here supports an interpretation that performance on the Sternberg task with supra-span memory sets was maintained because participants managed their search when challenged. To place this in context, if each word in the memory set was considered independently in a sequential self-terminating search, a participant would make multiple comparisons per probe word until they reached (or never reached) a target word to be recalled. On positive trials in which the word *could* be reached, the location of the target word in the sequence would be critical to participant speed/efficiency, and the number of non-target words (probes) still present after the target word was reached would be irrelevant. On negative trials the total number of comparisons should reflect the size of the memory set to be recalled and the number of probes to consider. Because, in all these conditions we did not observe a rise in reaction time, we conclude that the increased cognitive load in our experiments, whether through larger memory sets or the addition of a working memory task, precipitated participant adaptations to improve efficiency. Memory alone appears to have made a relatively small contribution to increased reaction time that was overshadowed by other processes. Applying Stemberg's (2016) logic, short term memory seems not to compete for resources with procedural nor meta-memory, the memory for the task itself, nor with perceptual and decision-making processes.

Study Limitations and Directions for Further Research

We can confidently eliminate, from our data, the possibility that the participants' slopes for reaction times were higher in response to large versus small memory sets. In interpreting our data, an understanding regarding the effects of other factors such as the varied interaction effect sizes is limited by the number of trials per participant and the number of participants in our study. Also, we embedded our research in a task that simulated a real-world task, and this may have produced a different pattern of results than would be found in laboratory tasks with more artificial stimuli. We did not investigate how reading behavior might have interacted with memory in the search process. And, as suggested by one reviewer, we did not analyze how participants' gender may have impacted these results. Of note in this regard, while our sample was predominantly female and the generalizability of these findings might be questioned, our random assignment of participants to conditions led to similar numbers of men in each condition, eliminating concerns about a gender confound in these results. Future studies might address each of these remaining concerns.

Conclusion

In three separate experiments of this study, we asked participants to find memorized target words in a list of probes. These memory sets varied in size, and task difficulty was controlled in several ways (the number of probes varied in size, the target position withing the probes was varied and in the last experiment a working memory task was added). We found participants' management of these challenges to be surprisingly efficient and accurate. Although there was a main effect of increased memory set size on slower reaction times, we found per-item memory efficiency to be as good or better for larger versus smaller memory sets. This somewhat surprising discovery differs from previous findings when fewer probe items were used (Okada & Burrows, 1978), challenging these prior investigators' earlier supposition that memory difficulty might increase with the logarithm of memory set size.

Acknowledgments

We would like to thank Julie Sabourin, Eryn Wright, Geneviève Taylor and Catherine Levac for developing stimuli and running experiments.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Laurentian University Research Fund grant to the first author.

ORCID iD

Josée Turcotte https://orcid.org/0000-0002-4234-3680

References

Barberis, N., Quattropani, M. C., & Cuzzocrea, F. (2019). Relationship between motivation, adherence to diet, anxiety symptoms, depression symptoms and quality of life in individuals with celiac disease. *Journal of Psychosomatic Research*, *124*(1), 1–6. https://doi.org/10.1016/j.jpsychores.2019.109787

Brigg, G. E. (1974). On the predictor variable for choice reaction time. *Memory and Cognition*, *2*(3), 575–580. https://doi.org/10.3758/BF03196923

Burrows, D., & Okada, R. (1975). Memory retrieval from long and short lists. *Science*, *188*(4192), 1031–1033. https://doi.org/10.1126/science.188.4192.1031

Corbin, L., & Marquer, J. (2009). Individual differences in Sternberg's memory scanning task. *Acta Psychologica*, *131*(2), 153–162. https://doi-org.librweb.laurentian.ca/10.1016/j.actpsy.2009.04.001

Corbin, L., & Marquer, J. (2013). Is Sternberg's memory scanning task really a short-term memory task? *Swiss Journal of Psychology*, *72*(4), 181–196. https://doi.org/10.1024/1421-0185/a000112

Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, *94*(446), 496–509. https://doi.org/10.1080/01621459.1999.10474144

Johns, E. E., & Mewhort, D. J. K. (2002). What information underlies correct rejections in short-term recognition memory? *Memory & Cognition*, *30*(1), 46–59. https://doi.org/10.3758/BF03195264

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*(6), 609–626. https://doi.org/10.1037/0033-295X.89.6.609

Murdock, B. B. (1983). A distributed memory model for serial-order information. *Psychological Review*, *90*(4), 316–338. https://doi.org/10.1037/0033-295X.90.4.316

Murdock, B. B. (1995). Developing TODAM: Three models for serial-order information. *Memory & Cognition*, *23*(5), 631–645. https://doi.org/10.3758/BF03197264

Okada, R., & Burrows, D. (1978). The effects of subsidiary tasks on memory retrieval from long and short lists. *The Quarterly Journal of Experimental Psychology*, *30*(2), 221–233. https://doi.org/10.1080/14640747808400671

Rubio-Tapia, A., Hill, I. D., Kelly, C. P., Calderwood, A. H., & Murray, J. A. (2013). ACG clinical guidelines: Diagnosis and management of celiac disease. *American Journal Gastroenterology*, *108*(5), 656–676. https://doi.org/10.1038/ajg.2013.79

Shiffrin, R. M., & Raaijmakers, J. (1992). The SAM retrieval model: A retrospective and prospective. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *Essays in honor of William K. Estes, Vol. 1. From learning theory to connectionist theory; Vol. 2. From learning processes to cognitive processes* (pp. 69–86). Lawrence Erlbaum Associates, Inc.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166. https://doi.org/10.3758/BF03209391

StataCorp (2019). *Stata statistical software: Release 16.1*. StataCorp LLC.

Sternberg, S. (2016). In defense of high-speed memory scanning. *The Quarterly Journal of Experimental Psychology*, *69*(10), 2020–2075. http://dx.doi.org/10.1080/17470218.2016.1198820

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7