THE EFFECT OF LOOKING UP INFORMATION ON LEARNING AND LONG-TERM MEMORY

BY

GEOFFREY L. MCKINLEY

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Aaron S. Benjamin, Chair
Professor Gary S. Dell
Associate Professor H Chad Lane
Professor Daniel G. Morrow
Associate Professor Lili Sahakyan

Abstract

Retrieval is a potent method of learning, with a variety of indirect and direct benefits. The *testing effect* describes the finding that retrieving information enhances long-term retention of that information, relative to restudying. Learners appear to be unaware of this benefit, and in turn, underutilize retrieval. As technology has made a vast amount of information more accessible, it has created an environment that disincentivizes retrieval from memory. The current studies examine the relationship between lookup behavior, and later memory for the material. Chapter 2 examines how the imposition of an external access cost influences lookup behavior and memory, when such behavior is disincentivized by making the act of looking information up more perceptually difficult (Experiment 2), or by decreasing the responsiveness of the lookup device (Experiment 1, 3, and 4). Chapter 3 examines the role that memory self-efficacy has on lookup behavior and memory. In Experiment 5, participants are given false feedback about their performance on a prior memory test. In additional analyses based on the data from Experiments 3 and 4, I directly correlate measures of memory self-efficacy with lookup behavior and memory. An access cost reduced lookup behavior, independent of the type of cost (Experiments 1 and 2), but did not affect lookup behavior when the access cost was manipulated within-subjects (Experiment 4), or when the size of the cost was unpredictable (Experiment 3). Importantly, in all of the conditions in which lookup behavior was successfully discouraged by an access cost, memory was enhanced. More lookup behavior was associated with better memory, even after controlling for memory self-efficacy (Experiment 3 and 4). In addition, higher memory self-efficacy was strongly associated with memory (Experiment 3), a result that is likely due to the greater influence of memory self-efficacy on tasks in which participants have more control over their learning. The difficulty of manipulating memory self-efficacy and lookup behavior is discussed.

To My Father

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

The advent of the smart phone has provided people with the opportunity to access any public information from the World Wide Web at nearly any time. The value that this service has for society is obvious: it provides opportunities for people to access information, in real-time, from an enormous bank of knowledge, to serve our immediate goals. When one wants to access some bit of information, this technology gives one the choice between querying their memory and querying the Internet. Given the efficiency of smart phone technology, and the fallible nature of human memory, it would seem reasonable to assume that many would choose to look this information up on the Internet, rather than relying on their own memory. Although this choice may serve one's immediate goals, one potential consequence is that it discourages retrieval from memory, which may have negative long-term consequences for retention of that information.

Questions about the tradeoff between lookup behavior and retrieval from memory are directly relevant to the recent surge of interest in understanding the role of technology on our behavior and memory (Henkel, 2014; Sparrow, Liu, & Wegner, 2011; Storm & Stone, 2015; Storm, Stone, & Benjamin, 2017; Ward, 2013). For example, Sparrow et al. (2011) asked participants easy or hard trivia questions followed by a modified Stroop task in which they indicated whether the color of a word was blue or red. Critically, the word was either related to computers and search engines (e.g., "browser" and "Google"), or not (e.g., "book" and "Nike"). They found that the color identification task was performed slower when the words were related to technology. In addition, this effect was much larger after questions that could not be answered. The interpretation of this finding was that while trying to answer trivia questions, technology-related words became activated, which in turn, interfered with the color identification task. The results further suggest that, during the difficult trivia questions, technology-related words became more activated, perhaps because they provide a more desirable means of accessing the correct answer.

Other studies have found that participants tend to recall information better when they believed that it would be erased from digital storage (Sparrow et al., 2011), encode new information more effectively when they believe that the previous information would be saved (Storm & Stone, 2015), and show poorer recognition for objects that one took a picture of (Henkel, 2014). These studies all focus on the effect of technology during initial encoding. As such, they neglect one of the most potent tools for learning: retrieval.

<u>The cognitive benefits of testing</u>

Testing oneself on material has many practical benefits (see Benjamin & Pashler, 2015; Roediger, Putnam, & Smith, 2011). The most broadly appreciated benefit is that it provides educators with diagnostic information about a student’s level of understanding or mastery of the material. However, testing also provides a learner with more direct benefits, as it improves their learning and retention. The mnemonic benefit that testing provides has been known about for many years (e.g., Glover, 1989; Spitzer, 1939); however, only in the last few decades has this effect received broad attention among memory researchers. There is now a greater focus on the functional role that testing has in the learning process. This conclusion stems primarily from research on what has been dubbed “the testing effect,” which refers to the advantage in memory that follows testing relative to restudying information (Roediger & Karpicke, 2006). The testing effect has been observed across many different stimuli, such as verbal material, including words (Carpenter & DeLosh, 2006) and paired associates (Jacoby, 1978; Karpicke & Roedgier, 2007a, Pyc & Rawson, 2010), name learning (Carpenter & DeLosh, 2005; Landauer & Bjork, 1978), as well as nonverbal materials such as map learning (Carpenter & Kelly, 2012; Carpenter &

Pashler, 2007). More importantly, the effect has been observed with more meaningful stimuli, such as text passages (Roediger & Karpicke, 2006; see also Glover, 1989) and foreign language learning (Carpenter, Pashler, Wixted, & Vul, 2008; Carrier & Pashler, 1992; Karpicke & Roediger, 2008; Toppino & Cohen, 2009).

Aside from the direct benefits of retrieval, there are also more indirect benefits of retrieval (Roediger & Karpicke, 2006). For example, testing has been shown to improve later learning. More specifically, retrieval seems to facilitate retention of later information and reduce interference from proximal study lists (e.g., Szpunar, McDermott, & Roediger, 2008), even when participants retrieve unrelated information (Divis & Benjamin, 2014; Pastötter, Schicker, Niederhuber, & Bäuml, 2011). Retrieval has also been shown to benefit higher order cognitive skills, such as generalization (McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013; see also Chan, McDermott, & Roediger, 2006; Jacoby, Wahlheim, & Coane, 2010) and category learning (Jacoby et al., 2010), as well as semantic organization, inference, and concept learning (Karpicke & Blunt, 2011).

Testing can also provide useful metacognitive information. Learners can use tests to assess their own state of learning for individual materials (Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991). Similarly, testing can have downstream effects in which testing provides more accurate information about the difficulty of material, leading to more effective learning on subsequent study sessions with different material (Benjamin, 2003; Keleman, Winningham, & Weaver, 2007). Such opportunities can also provide learners with information about the effectiveness of a given study strategy, which can also give some indication to learners about which strategies are most effective (Finley & Benjamin, 2012; Metcalfe, Kornell, & Son, 2007).

Metacognitive appreciation of the benefits of testing

Despite the many desirable consequences of retrieval, learners appear to be somewhat unaware of the advantages of testing (Karpicke, 2009; Karpicke & Blunt, 2011; Karpicke & Roediger, 2008; but see Tullis, Finley, & Benjamin, 2013; Tullis, Fiechter, & Benjamin, 2018). In one survey, only 42% of participants reported that they would rather test themselves on the material that they just learned, and the overwhelming majority of those participants (at least 72%) stated that they self-tested for its diagnostic, rather than mnemonic, value (Karpicke, Butler, & Roediger, 2009, see also Kornell and Bjork, 2007). This is most puzzling when one considers the fact that students picked this option even in situations that permitted restudy after practice tests. This bias is also evident in laboratory experiments. For example, in Karpicke (2009), participants were given paired associates to study over multiple study-test cycles. Once an item was successfully retrieved, items were either given additional testing, studying, or were dropped from both study and test. Items that were given additional testing were substantially more likely to be remembered on a final test than were items that received additional study (or the items that were dropped). However, items that were given additional study received higher judgments of learning (JOLs) than items that were given additional testing—in fact, JOLs for additionally-tested items were rated at about the same level as those that were dropped.

Unsurprisingly, these misleading metacognitive beliefs are also reflected in how learners choose to study. When participants in Karpicke (2009) were given control over their study, their choices to restudy, test, or drop were consistent with their JOLs. That is, participants seemed to drop items that they gave the highest JOLs to, test themselves on items that were given lower JOLs, and restudy items that they gave the lowest JOLs to. It should be noted that although there is some evidence that participants' choices to restudy are sometimes driven by rational considerations of item difficulty, they nevertheless appear to under-utilize testing in their study regimes (Tullis et al., 2018). This is most evident in the fact that participants choose to restudy, even when feedback after the practice test will be given or additional study opportunities are available.

<u>The tradeoff between retrieval and lookup behavior</u>

Taken together, these findings indicate that participants are unaware of the benefits of retrieval and consequently, they scarcely use this very potent method for enhancing memory. Given this general aversion to using retrieval, it is reasonable to think that having the option to look up information would likely exacerbate this pre-existing tendency to avoid using retrieval. Indeed, there is some research on strategy use that supports this claim, much of it using a simple paired-associate learning task. Many of the studies using this task were concerned with older adults, and how their behavior and memory differs from young adults. Therefore, although much of the discussion below describes this work with adults across the lifespan, the research in this dissertation is motivated by generalizing some of these ideas and methods to younger adults.

Ackerman and Woltz (1994) provided an early demonstration of how a lookup task could work. Many of the relevant studies on strategy use employ some variation of this noun-pair lookup task, and a variation will be used in the current research, so this task will be carefully described in what follows. In this task, participants are shown a grid of words on the upper half of the computer screen. A series of words are placed in the first row of the grid with another series of words placed directly below each word in the first row. This constitutes the first set of noun pairs in what will be referred to as a "key." On each trial, a noun pair is presented on the bottom center of the screen, referred to as the "probe," and the participants' task is to simply indicate whether or not the pair is also present in the key. On all of the trials, the probe consists of one word from the top row, and another from the bottom row. On half of the trials, both of these words are in the same column in the key. These are called positive trials because the participant must respond "Yes" to be correct. On the other half of the trials, the words that make up the probe are in different columns in the key. These are referred to as negative trials because the participant must respond "No" to be correct. In a consistent-mapping (CM) condition, the words are randomly arranged in the key with the constraint that each word in a noun pair is in the same column as its counterpart and that words do not change rows. In a varied-mapping (VM) condition, the words are randomly arranged with the only constraint that words do not change rows. Typically, participants are given many trials for each condition, and the same noun pair will often be tested multiple times throughout the experiment. Because the noun pairings remain consistent in the CM condition, participants will eventually learn those pairings. However, in the VM condition, where the nouns are constantly being re-paired, the participant will have to look up the current pairings to make each response.

The value of this task primarily lies in the fact that because the words in the VM condition are not consistently paired, there is no opportunity for participants to shift from a reliance on the key to a reliance on memory. However, in the CM condition, there is opportunity to learn. Therefore, in the VM condition one must always scan the key in order to perform above chance. In contrast, the CM condition allows one to use their memory in performing the task, though, notably, it is not required. As a result, because VM trials require scanning, the response times (RTs) that come from these trials index the amount of time it takes for a participant to search the key. One can use these RTs as a baseline measure for scanning behavior to roughly infer whether a participant is relying on memory or scanning the key in the CM condition. Not surprisingly, early CM trials yield RTs that look indistinguishable from those from VM trials, but they decrease as participants rely more on retrieval in the CM conditions. More succinctly, participants rely more on retrieval as the task progresses, and the shift to retrieval is revealed by the point in which CM trials begin to be completed faster than VM trials.

Using this task, Ackerman and Woltz (1994) found that although many participants shifted from a scanning strategy to a retrieval strategy, some persisted in scanning despite the inefficiency of that choice. The participants who failed to shift tended to score lower on measures of perceptual speed and reasoning ability, indicating that the shift to retrieval may be partly due to differences in learning between participants. Indeed, Rogers, Hertzog, and Fisk (2000) found that older adults were more reliant on scanning behavior than young adults, and shifted to retrieval later than did young adults, a finding that is quite common in this literature (e.g., Rogers & Gilbert, 1997; Touron & Hertzog, 2004a, 2004b). Further, they found that older scanners performed more poorly on measures of ability, including working memory, associative learning, semantic memory access, and perceptual speed, compared to older retrievers. Results of this kind have suggested that the reason older adults show this delayed shift to retrieval is that their memory for the pairs is insufficient to support the shift. This sort of "bottom-up" explanation implies that this delayed shift is observed in older adults because of age-related differences in learning and memory (Touron, Hoyer, & Cerella, 2001, 2004).

However, other results have suggested that older adults may avoid retrieval for reasons that are unrelated to memory ability. This "top-down" explanation has been referred to as the *retrieval aversion hypothesis* (Touron & Hertzog, 2004a; Rawson & Touron, 2009), and states that older adults are more averse to using retrieval in general, independent of how well they have actually learned the information. This bias could be due to a variety of factors, such as lower confidence in their memory ability (Touron & Hertzog, 2004b; see also Hertzog & Hultsch, 2000), a more conservative criterion for using retrieval (cf. Fraundorf, Hourihan, Peters, & Benjamin, 2018), a greater emphasis on accuracy, or the greater effort older adults must put into retrieval (see Touron & Hertzog, 2004a). For example, in Touron and Hertzog (2004a), young and older adults participated in a noun-pair lookup task, in which half of the participants were given additional interspersed trials (for half of the noun pairs) without the key present. They reasoned that if older adults used retrieval less because of slower learning, then they should use retrieval on a standard lookup trial when they had correctly responded to the preceding memory-probe trial for that item. That is, the conditional probability of using retrieval on a standard lookup trial, given that they had correctly recognized the pair on a recent memory-probe trial, was used as an indirect index of retrieval avoidance. If the retrieval aversion hypothesis is correct, this measure would be lower for older adults than for young adults. Indeed, this measure was lower, and increased more slowly over trials for older adults than for young adults.

In another experiment, participants learned either half or all of the noun pairs to criterion before the noun-pair lookup task. According to the retrieval aversion hypothesis, they reasoned that older adults would still show a delayed strategy shift for prelearned noun pairs. They found that although prelearning all of the pairs before the noun-pair lookup task substantially reduced RTs for older adults, it did not eliminate differences in their use of retrieval, as older adults reported using the lookup key even for noun pairs that had been prelearned (see also, Hines, Hertzog, & Touron, 2012).

The results above suggest that older adults are choosing not to retrieve early on in the process upon seeing a given probe. In other words, it is possible that older adults are prematurely selecting a strategy when trying to respond to a probe. Indeed, there is some evidence that strategy selection does precede answering questions in young adults. In one example, Reder (1982) had participants read stories and make plausibility judgments or recognition judgments about additional sentences either immediately after, twenty minutes after, or two days after reading the story. She reasoned that if strategy is not selected beforehand, then plausibility judgments should not be affected by the interval between reading the story and making these judgments. However, she found that plausibility judgments for not-presented sentences took longer to make when made immediately after reading the story. Based on this result, she argued that participants in the immediate condition might "have tried to answer questions by matching word for word", and that because the sentence had not been presented, the plausibility judgment took longer to make. In other words, this result suggests that participants were more likely to try and search their memory to make a plausibility judgement, but only when they had been exposed to the information recently.

In another study (Reder, 1987), participants were shown a series of general knowledge questions. Half of the participants were required to make speeded yes/no judgments as to whether they knew that answer or not. If they responded "Yes," they were required to retrieve the answer. The other half of the participants were instructed to answer the question as quickly as possible, if they knew the answer. She found that participants who made a yes/no judgment were faster to make this response, even for positive responses that were followed by the answer. This finding was interpreted as evidence that participants can assess their ability to retrieve an answer before actually doing do so (cf. Benjamin, 2005). Thus, there is some evidence supporting the notion that strategy selection can precede access of more probative target information.

Many of the conclusions about retrieval avoidance may be useful in understanding behavior and memory in adults more generally. For example, Touron and Hertzog (2004b) manipulated the key in a lookup task in a way that varied in terms of perceptual and memory load. The key was populated either by a small or large number of noun pairs, which characterized a low or high perceptual load, respectively. To manipulate memory load, the noun pairs on the screen were either from a small or large pool of noun pairs, which characterized a low and high memory load, respectively. They found that participants shifted to using retrieval (via self-reported strategy use) earlier when the perceptual load was high, and the memory load was low, compared to when both the memory and perceptual load were low. That is, when the scanning strategy was more effortful, participants switched to retrieval sooner than when the scanning strategy was less effortful. That fact—that the nature of the lookup interface influences choices about retrieval—is central to the first few experiments of this dissertation.

To summarize, there are many benefits of retrieval for learning and memory, yet learners seem to be relatively unaware of these benefits, and therefore, appear to under-utilize testing in their study habits. Some learners avoid using retrieval, even after the materials are sufficiently well learned to do so. This bias may be due to retrieval being too effortful, to an undue emphasis on accuracy, or to lower confidence in one's own memory. Yet, retrieval can be encouraged by increasing the relative cost of using a lookup device. Taken together, the implications of these findings suggest a potentially deleterious effect of having readily access to the Internet.

<u>Interface design and lookup behavior</u>

There is some research about how user interfaces influence behavior. Much of this research is motivated by the *soft constraints theory* (Gray & Fu, 2004), which provides an account of interactive behavior in the context of human-computer interaction. It states that many tasks possess "hard constraints" and "soft constraints" which influence how users will complete a given task. While hard constraints mandate the range of possible behaviors, soft constraints serve to bias certain behaviors. For example, to call a friend on one's own cell phone, one can retrieve the number from memory and enter the number in the keypad, or search for their friend in their contacts. Although the task can be done either way, the fact that the number can be accessed on the phone is a soft constraint that will likely discourage any commitment of that number to memory. Thus, cognitive strategies are adapted to the constraints of the task environment, or in this case, the accessibility of the phone number.

One common way of studying the effects of soft constraints on interactive behavior uses a copying task called Blocks World Task (BWT). In this task, participants are required to drag colors into a grid so that it matches the ordering of the colors in a target grid next to it. Critically, to manipulate access cost, the target grid is either visible for the entire trial (Low cost), or covered by a mask that can be uncovered immediately after clicking on the target grid (Medium cost), or after a delay after clicking (High cost). Using this general method, some researchers have found that increasing the access cost induces a more memory-based strategy (e.g., Fu & Gray, 2000; Gray, Sims, Fu, & Schoelles, 2006; Morgan, Patrick, & Tiley, 2013; Patrick et al., 2015; Waldron, Patrick, & Morgan, 2007; see also Gray & Fu, 2004). This shift is indicated by several effects, including a reduction in the number of times the target window is accessed, and an increase in the number of colors that are correctly placed before revisiting the target window. Some have directly tested memory and found that imposing this high access cost can enhance performance on memory tasks (e.g., Morgan, et al., 2009; Waldron et al., 2007). That is, when participants are asked to recreate the grid after a given trial, access cost was positively related to the number of colors that were correctly placed. However, it should be noted that this strategy comes with a few costs, such as longer visit times for the target window, longer completion times, as well as an increase in error rates (though the error rates are quite low in absolute terms).

It has also been shown that the memory-based strategy that is induced by this higher access cost can also protect against the negative effects of interruption on performance (Morgan & Patrick, 2013; Morgan et al., 2013; Morgan et al., 2009). This is often indicated (after interruption) by a greater number of trials that are completed without revisiting the target window, an increase in the number of colors that are correctly placed before revisiting the target window, as well as a shorter resumption lag. This latter benefit is also found for trials in which there is no opportunity to revisit the target window after interruption (Morgan, et al., 2009, Exp. 3). That is, participants in the high access cost condition show better memory for behavior that was planned before an interruption (see also Morgan & Patrick, 2013; Waldron, Patrick, & Duggan, 2011).

Storm, Stone, and Benjamin (2017) reported a set of experiments in which people face a choice between retrieving information and looking it up. In their experiments, participants were given a set of difficult trivia questions to answer. One group of participants was required to rely on memory to do so; another was required to use the Internet to get the answer. In a second phase, all participants were given a new set of easy trivia questions, and were given the option to either rely on their memory or to use the Internet to answer them. Critically, using the Internet during the second phase was made to be less convenient for some of the participants. Participants who could use the Internet during the first phase were more likely to use the Internet during the second phase than those who were not allowed to use the Internet. More importantly for present purposes, participants were also less likely to use the Internet when access to the lookup device was inconvenient.

Taken together, the above research suggests that retrieval can be encouraged or discouraged by manipulating the ease of looking up the information. Allowing one to have ready access to information will decrease the probability that they will try to retrieve that information from memory. Conversely, making it more difficult or time-consuming to look information up will increase the probability that one will use retrieval. It remains unclear, though, how these manipulations will affect memory, as few studies have directly examined this question. There have been studies that have measured memory after completing (or during) the noun-pair lookup task (e.g., Rogers et al., 2000). However, most of this research focused on the mnemonic consequences of scanning or retrieving, rather than trying to actually manipulate lookup behavior. In examining whether perceptual and memory load affected lookup behavior, Touron and Hertzog (2004b) measured memory on an associative recognition test. They did not find any reliable effects, though, it should be noted that performance was very high (average *d*'s > 3.6), indicating the possibility of ceiling effects obscuring any memory benefits. As described above using the BWT, there is some evidence that reliance on a display-based strategy decreases memory in research, as evident in the number of colors that are correctly placed during a test trial; however, there are some notable differences between this task and the noun-pair lookup task.

One obvious difference is that the BWT relies on encoding visual information such as color and position, whereas the lookup task uses verbal information. Also, in the BWT, each time a participant accesses the target window, they can control the duration of the target window. Therefore, such effects are likely to be driven in part by differences in study time. Most importantly, the memory tests of the BWT occurred immediately after completing the trial (Morgan et al., 2009, Exp.1), or after an interruption delay of only five seconds (Morgan et al., 2009, Exp. 3). These results thus do not speak to the long-term consequences of lookup versus retrieval. Further, when a trial is interrupted, and then tested after a delay, performance is not just driven by memory for the target grid, but also by ones' memory for unfinished goals that were planned before the interruption.

The purpose of the current experiments is to examine the consequences of looking up information on long-term memory and learning. In the experiments presented here, I examine how characteristics of the lookup interface and of the participants themselves influence ones' decision to retrieve or look something up and consequent memory. A modified version of the noun-pair lookup task will be used in all of the current experiments, with small variations across experiments. In Chapter 2, we manipulate the ease of looking information up and examine its effect on lookup behavior and subsequent memory. In Experiment 1, the cost of looking information up is manipulated by increasing the delay between clicking on the current cue and receiving the sought-after target word. In Experiment 2, the cost of looking up is manipulated by increasing the perceptual density of the search display for finding the cue. In Experiment 3, the design from Experiment 1 is modified in an attempt to make the manipulation stronger. I do so by making the time cost of accessing information longer and variable in the costly condition. Finally, in Experiment 4, the cost of accessing information is manipulated within-subjects.

In Chapter 3 I explore the effect of memory self-efficacy in one's memory on lookup behavior and memory. This is done by providing false feedback on a memory test prior to the lookup task (Experiment 5). In this chapter, I also present questionnaire data from Experiments 3 and 4 to examine the role of memory self-efficacy on lookup behavior and memory. Using mediation models, I try to hone in on the relationship between lookup and memory, while excluding the direct effects of memory self-efficacy on these variables. I also examine whether lookup behavior mediates the relationship between memory self-efficacy and memory using mediation models. The Appendix to the dissertation includes analyses of gender differences in lookup behavior and memory for Experiments 1, 2, and 3, and a brief discussion of those effects.

All of these studies use a variation of the noun-pair lookup task. One aspect of the noun-pair task used in much of the above research is that it relies on response times to infer the strategy that is being used. Although there is evidence that response times are valid indicators of strategy (e.g., Touron, Hertzog, & Frank, 2011), I take a different approach here. In the experiments presented below, participants were required to click on the cue in order to see its corresponding target. Therefore, lookup behavior was directly measured. The probe on each trial is a cue, and the appropriate response is the matched target term. To report this term, learners had to either retrieve it from memory or access it by clicking on the appropriate box in the key. The positions of the noun pairs in the key were randomly arranged on each trial. In each block, each cue served as a probe once. The order in which the cues were tested was random with the constraint that each cue served as the probe only once in each block. From a participant's perspective, there was no indication of when a block started or finished.

CHAPTER 2: THE EFFECT OF USER INTERFACE ON RETRIEVAL USE

Experiment 1: Effect of access cost on retrieval

The purpose of this experiment was to examine whether making the act of looking information up more time-consuming would discourage lookup behavior, and whether this reduction in behavior would have positive consequences for memory.

Method

Subjects. 155 introductory level psychology students participated in this experiment for partial course credit. However, 5 of the participants did not return for the second part of the experiment, yielding 150 participants.

Design. This study followed a mixed 5 (Block) x 2 (Cost: cheap vs. costly) design. Each participant experienced 5 blocks of the look-up task (within-subjects). The cost of looking information up was manipulated between-subjects. Half of the participants were assigned to a "costly" condition, in which access to the target was briefly delayed, and the other half were assigned to a "cheap" condition, in which access to the target was immediate.

Materials. A list of paired associates was created using a list of 40 words. Each word was randomly assigned to be a cue or a target, and each target was randomly paired with a cue, yielding a list of 20 paired associates. Words were specifically chosen to be of moderate frequency (M = 8.91, SD = 1.07, Min = 7.51, Max = 11.59), and between 7 and 9 letters in length. These paired associates were the same across participants.

Procedure. Participants completed the lookup task as described above. The key consisted of 20 nouns; each one served as a probe once in a given block. There was a total of 5 blocks of 20 trials each. In the costly condition, when participants clicked on a cue in the key, the target appeared after a 1.5 second delay. In the cheap condition, there was no such delay.

After completing the lookup task, participants were excused for the day. They returned the next day and were given a cued recall test on the paired associates from the lookup task. The order in which each pair was tested was randomized for each participant. An example trial of this task is shown in Figure 1.

Results

Lookup behavior. The data are shown in Figure 2, which shows lookup rate by block and condition. A 5 (Block) x 2 (Cost) Mixed ANOVA was conducted on the lookup behavior. This analysis revealed a significant effect of Block, $F(4, 592) = 411.84$, $p < .05$, indicating that the number of lookups decreased across Block. There was a tendency for participants in the cheap condition to exhibit more lookup behavior than the costly condition ($M = 3.70$, $SD = 0.84$ and $M = 3.42$, $SD = 0.96$, respectively), though this was not significant, $F(1, 148) = 3.45$, $p > .05$. The Block x Cost interaction was also not significant, $F(4, 592) = 1.99$, $p > .05$. Collapsed across Block, lookup behavior did not differ between the cheap and the costly condition, $t(146.13) = 1.86$, $p > .05$, $B_{01} = 1.17$, $d = 0.30$.[1]

Memory. Memory for the costly condition ($M = 0.57$, $SD = 0.23$) was numerically higher than memory for the cheap condition ($M = 0.51$, $SD = 0.27$), though this was not significant, $t(142.78) = -1.53$, $p > .05$, $B_{01} = 1.95$, $d = 0.25$. These data are shown in Figure 3.

Discussion

Although there was a tendency for those in the costly condition to look up information less frequently, this effect was not significant. Similarly, although participants in costly condition remembered more than those in cheap condition, this difference was not significant. It should be emphasized that the difference in memory is expected to be dependent on the

[1] A Welch's t-test was conducted for all between-subjects t-tests.

difference in lookup behavior. Therefore, it is possible that the manipulation was not strong enough to encourage sufficient reliance on retrieval. Experiment 2 utilizes a very similar design as Experiment 1 but manipulates access cost in terms of time and effort.

Experiment 2: Effect of perceptual cost on retrieval

In Experiment 1, the results were numerically consistent with our predictions, despite the fact that the manipulation consisted of a short time delay in the costly condition. Therefore, it is possible that this manipulation was not strong enough to effectively discourage lookup behavior. The purpose of this experiment was to examine whether the cost of access would generalize to another form of access cost: perceptual effort. It should be noted, though, that this manipulation was expected to also represent a larger cost also in terms of time, as the greater perceptual effort would also likely be associated with a greater time cost.

Method

Subjects. A total of 134 introductory-level psychology students participated in this experiment for partial course credit. Five participants did not return for the second part, yielding a total of 129 participants.

Design. This study followed a mixed 5 (Block) x 2 (Cost: cheap vs. costly) design. Each participant experienced 5 blocks of the lookup task (within-subjects). The cost of looking information up was manipulated between-subjects. Half of the participants were assigned to a costly condition, in which the search grid was dense, and the other half were assigned to a cheap condition, in which the search grid was sparse.

Materials. The words in this experiment were identical to the words in Experiment 1. However, in the costly condition, an extra 22 Lorem Ipsum words (i.e., nonwords) were included as distractors for the lookup task. These words did not change across participants.

Procedure. The cheap condition was identical to the cheap condition in Experiment 1. In the costly condition, there was an additional 22 nonwords that served as distractors. The display thus included 42 boxes, only 20 of which were relevant for the task. This greater grid density made searching for the cues more difficult. In both conditions, the target appeared after no delay once the appropriate box was pressed. If a participant clicked on a nonword, the box was highlighted but no word appeared. Participants would then be told that they clicked on the incorrect cue. None of the nonwords served as probes. As in Experiment 1, there was 5 blocks in the lookup task. After completing all 5 blocks, participants were excused for the day. They returned the next day and were given a cued recall test on the paired associates from the lookup task. The order in which each pair was tested was randomized for each participant. The nonwords were not tested.

Results

Lookup behavior. A 5 (Block) x 2 (Cost) Mixed ANOVA revealed a main effect of Block, $F(4, 508) = 300.08$, $p < .05$, indicating a tendency to look up information less often in subsequent blocks. There was a tendency for participants in the cheap condition to exhibit more lookup behavior than in the costly condition (M = 3.82, SD = 0.90 and M = 3.62, SD = 0.85, respectively), but this was not significant, $F(1,127) = 1.637$, $p > .05$. The Block x Cost interaction was also not significant, $F(4, 508) = 1.193$, $p > .05$. Collapsed across Block, lookup behavior did not differ between the cheap (M = 3.82, SD = 0.90) and costly condition (M = 3.61, SD = 0.85), $t(126.37) = 1.29$, $p > .05$, $B_{01} = 2.49$, $d = 0.23$, though the pattern was numerically consistent with the predictions. These data are shown in Figure 4.

Memory. Cued recall performance was numerically worse for the cheap condition (M = 0.47, SD = 0.29) than for the costly condition (M = 0.54, SD = 0.27), though this was not significant, t(126.33) = -1.33, p > .05, B_{01} = 2.40, d = 0.23. These data are shown in Figure 5.

Discussion

Participants in the cheap condition tended to recall less information than participants in the costly condition; however, this was not significant. Similarly, although there was a tendency for those in the costly condition to look up information less often, this was also not significant. It appears, then, that a larger set size did not have a considerable effect on lookup behavior.

The finding that more perceptual difficulty did not have a substantial effect on lookup behavior is puzzling because not only was it more difficult to find the correct cue in the costly condition, but it also took longer. For example, in the first block—in which participants had to look up every item—participants in the cheap condition took 3.73 seconds before clicking on the correct cue, whereas those in the costly condition took 7.04 seconds before clicking on the correct cue, t(105.52) = -15.68, p < .05, $B_{10} = 3.03 \times 10^{28}$, d = 2.75. Therefore, it took almost 4 seconds more for the costly condition to click on the correct cue, compared to the cheap condition. In Experiment 1, the cost of looking information up had only a time component, whereas in Experiment 2, the cost included components of time and effort, and the time cost was greater than in Experiment 1 (at least in Block 1). Therefore, there are reasons to expect that the effect would have been larger in the second experiment. However, it is possible that the unfilled delay in Experiment 1 may actually be more frustrating since it is out of the participants' control.

Combined analysis of Experiments 1 and 2

In order to evaluate the effects in Experiments 1 and 2 with greater power, the data were collapsed across experiments, yielding 279 participants. A 5 (Block) x 2 (Cost) Mixed ANOVA revealed a main effect of Block, $F(4, 1108) = 712.19$, $p < .05$, indicating a reduction in lookup behavior across blocks. There was also a main effect of Cost, $F(1, 277) = 5.017$, $p < .05$, indicating less lookup behavior in the costly conditions than in the cheap conditions. There was also a Block x Cost interaction, $F(4, 1108) = 3.064$, $p < .05$, which stems from a larger effect of Block on lookup behavior for the costly conditions than the cheap conditions. Collapsed across Block, participants in the cheap conditions looked up the target word more often (M = 3.82, SD = 0.90) than those in the costly conditions (M = 3.62, SD = 0.85), $t(276.67) = 2.25$, $p < .05$, $B_{10} = 1.44$, $d = 0.27$. Most importantly, participants in the costly conditions (M = 0.56, SD = 0.25) remembered more on the cued recall test than participants in the cheap conditions (M = 0.49, SD = 0.28), $t(272.32) = -2.02$, $p < .05$, $B_{01} = 1.09$, $d = 0.24$.

A mixed-effects logistic model was conducted on the combined data in attempt to partial out item variability, and thus obtain a purer effect of the manipulation. More specifically, memory was modelled on a trial level, with random intercepts for participants and items using the `lme4` software package in R. The following equation was used:

$$(1)\ \text{logit}\,(P(Y_{ijk})) = \gamma_{jk} + \tau X_{ijk}$$

where $\gamma_{jk} = \gamma_{00} + u_{j0} + v_{0k}$, i denotes the trial, j denotes the participant, and k denotes the item, γ_{00} denotes the overall intercept, u_{j0} denotes the deviation of j^{th} participant's intercept from the overall intercept, v_{0k} denotes the deviation of the k^{th} item's intercept from the overall intercept, X indicates the Cost condition (costly = 1, cheap = 0), and Y indicates whether or not the item was recalled at test (1,0). The Cost variable was centered before entering it into the model. Table 1 shows the parameter estimates derived from the model. Importantly, there was a significant effect of Cost on memory ($b = 0.45$, SE = 0.20, $z = 2.235$, $p < .05$), indicating that the odds of recalling a target from the noun-pair lookup task was exp(0.45) = 1.57 times greater in the costly condition than in the cheap condition. Based on these results, I provisionally assert that these effects are real, but small in magnitude (i.e., d = 0.23) and thus requiring of either larger sample sizes or stronger manipulations.

Experiment 3: Effect of a greater and variable time cost on retrieval

In Experiments 1 and 2 there was a tendency for those in the costly condition to rely more on memory retrieval, and to recall more of the targets on a cued recall test. However, the effects were not significant in either individual experiment. As mentioned above, the time delay in Experiment 1 was quite short (i.e., 1.5 seconds), which may not have been a strong enough manipulation. Therefore, this experiment used a stronger manipulation of the access cost by making the access delay longer and more variable.

Method

Subjects. Data collection for this experiment is still ongoing. A total of 68 people have thus far participated in this experiment for partial course credit or $10 for compensation. However, one student had participated in a similar experiment and one did not show up for part 2, yielding a total of 66 participants. Random assignment yielded 33 participants in the cheap condition and 33 participants in the costly condition. These people were mostly students at the University of Illinois at Urbana-Champaign or recent graduates, but on some occasions, they were slightly older people from the community.

Design, materials, and procedure. Experiment 3 used the same design as Experiment 1 with the exception that the delay between clicking on a cue and the target appearing in the costly condition was randomly selected from a uniform distribution ranging from 1.5 to 4.5 seconds. In addition, after the memory test was completed, participants were given a questionnaire about their memory (see Chapter 3).

Results

Lookup behavior. A 5 (Block) x 2 (Cost) Mixed ANOVA was conducted on the lookup data, indicating a tendency to look up targets less often as a function of Block, $F(4, 256) = 149.25$, $p < .05$. There was no effect of Cost, $F(1,64) = 0.00$, $p > .05$, nor was there any Block x Cost interaction, $F(4, 256) = 0.09$, $p > .05$. These data are shown in Figure 6. Collapsed across Block, there was no effect of Cost, $t(63.98) = -0.02$, $p > .05$, $B_{01} = 3.96$, $d = 0.01$. Targets were looked up at the same rate in the costly condition (M = 3.51, SD = 1.01) than in the cheap condition (M = 3.51, SD = 1.03).

Memory. Participants in the cheap condition recalled slightly less (M = 0.55, SD = 0.33) than those in the costly condition (M = 0.60, SD = 0.29), though, this was not significant, $t(63.35) = -0.61$, $p > .05$, $d = 0.15$, $B_{01} = 3.37$. These data are showed in Figure 7.

Discussion

Participants in the cheap and costly condition looked up target words at roughly equal rates. Given this failure to induce different lookup rates, it is not surprising that performance on the cued recall test did not differ between conditions. Our achieved sample size is too low to reach solid conclusions, though the remaining planned participants would need to show an effect size of d = 0.37 on lookup behavior —much larger than the previous experiments—to yield a significant effect for the entire sample. It is for this reason that this experiment was temporarily abandoned in favor of the within-subjects version presented shortly (the required effect size for remaining subjects was d = 0.41 at the time of postponement). These results suggest that there was something about the costly condition in this experiment that undermined any effect of cost that was present in Experiment 1. Although it is possible that the increase in delay somehow encouraged more looking up behavior, it is not clear why this, by itself, would have this effect.

A more likely possibility is that the variability in the delay somehow interfered with encoding. Specifically, it may be that this variation reduced the predictability of the onset of the word, which resulted in poorer encoding of the associative pair. That is, perhaps because participants in the costly condition could not precisely anticipate when the target would appear, they had more trouble deploying attention to the word. This would imply that although participants in the costly condition were attempting to retrieve more often, the lack of onset predictability interfered with their learning. If this were so, we would still expect that participants in the costly condition would take more time before deciding to look at target up.

Indeed, of the four blocks in which participants needed to decide whether to look up a target word (i.e., not including the first block, in which participants had to look up every word), participants in the costly condition took longer to initiate lookup than in the cheap condition in 3 out of 4 of the blocks. Specifically, the average time to look up a target word in the cheap condition was 3.64, 3.85, 4.04, and 3.66 seconds, for blocks 2 to 5, respectively. In contrast, participants in the costly condition took an average of 3.85, 4.11, 4.04, and 4.15 seconds, respectively. Although these data are confounded by the different rates of looking up between conditions, it should be noted that the difference in the time to look up a target word was present in block 2. This is informative because not only was this the first block in which participants could decide to look something up, but more importantly, the rate of looking up target words was roughly equivalent between these two conditions in this block. Given the small effect size observed in the first two experiments, as well as the difficulty of finding a stronger manipulation, we decided to use a within-subjects design for Experiment 4.

Experiment 4: Effect of a within-subjects time cost manipulation on retrieval

In all of the experiments above, the cost of access was manipulated between-subjects. This choice was based on the concern of carry-over effects that may result in a within-subjects design (Patrick et al., 2015); however, the statistical benefit of using a within-subjects design may outweigh the cost of using a between-subjects design, and overwhelm such carry-over effects. Therefore, Experiment 4 was essentially a within-subjects version of Experiment 1.

Method

Subjects. A total of 96 people participated for partial course credit or $10 in compensation. These people were mostly students at the University of Illinois at Urbana-Champaign or recent graduates, but on some occasions, they were slightly older people from the community. Data from two participants were dropped because they did not show up for part 2. Data from another participant was also dropped because the person did not finish part 1 within the allotted time. This yielded 93 participants.

Design, materials, and procedure. A total of 28 associated pairs were used. This included the 20 pairs that were used in the previous experiments. For each participant, paired associates were randomly assigned to the cheap and costly condition. Each participant was shown two adjacent 7 x 2 cell grids, with a cue in each cell. One of the grids contained cells that had a red outline, and the other grid contained cells that had a blue outline. The target word would appear after a 1.5 second delay in one of the grids, and immediately after in the other grid. The assignment of color and grid (left or right) to condition was randomly assigned. An example trial is displayed in Figure 8. As done in Experiment 3, a cued recall test was given on the next day, and after the memory test was completed, participants were given a questionnaire about their memory (see Chapter 3).

Results

Lookup behavior. A 5 (Block) x 2 (Cost) ANOVA revealed a main effect of Block, $F(4, 368) = 199.80$, $p < .05$, indicating that lookup behavior decreased as a function of Block. There was no effect of Cost, $F(1, 92) = 0.06$, $p > .05$, nor was there a Block x Cost interaction, $F(4, 368) = 0.50$, $p > .05$. These data are shown in Figure 11. Collapsed across Block, there was no effect of Cost on lookup behavior, $t(92) = 0.64$, $B_{01} = 7.14$, $d = 0.03$, as the cheap condition ($M = 3.82$, $SD = 0.87$) looked up target words as often as the costly condition ($M = 3.79$, $SD = 0.87$). These data are displayed in Figure 9.

Memory. Participants in the cheap condition ($M = 0.55$, $SD = 0.28$) recalled target words at roughly the same rate as participants in the costly condition ($M = 0.56$, 0.29), $t(92) = -0.72$, $p > .05$, $B_{01} = 6.79$, $d = 0.05$. These data are shown in Figure 10.

Discussion

Despite the greater power that comes with a within-subjects design, it appears that access cost did not have any detectable effect on lookup behavior. Not surprisingly, then, performance on the cued recall test did not differ between the cheap and costly condition. There are a few potential explanations for this null effect. One possibility is that the trends in Experiments 1 and 2 were due to sampling error. Although this is possible, there is an abundance of prior research showing that access cost can influence lookup behavior, using the BWT, programming tasks, as well as with tasks that use the Internet to probe semantic knowledge. Although the noun-pair lookup task is somewhat different than these other tasks, Touron and Hertzog (2004b) used a similar task and did find that younger adults were much more likely to adopt a retrieval-based strategy when the perceptual load was larger.

Another possibility for why participants were not sensitive to the manipulation of cost is that perhaps the access delay was not long enough. Although some of the studies using the BWT had found an effect of access cost using a 1-second delay, delays closer to 2.5 seconds are reported more often in the literature. In fact, upon debriefing, some of the first few participants reported being surprised when they found out that some boxes had a longer delay than others. In response to this concern, participants were asked prior to debriefing whether there had been a delay between clicking a box and the appearance of the target word. If they reported that there was a delay, they were asked whether there was a delay for all of the boxes, or just some of them. Of the 86 participants who were asked these questions, only 44 of them reported that they noticed a difference in the delays. To test the idea that the delay was not long enough, the analyses were repeated on only the participants who noticed a delay. However, the results did not change. Those who noticed the difference in delay were just as likely to look up target words in the cheap condition (M = 3.81, SD = 0.85) as they were in the costly condition (M = 3.76, SD = 0.85), t(43) = 0.77, $p > .05$, $B_{01} = 4.58$, d = 0.06. Not surprisingly, these participants also recalled roughly the same proportion of words from the cheap condition (M = 0.57, SD = 0.27) as the costly condition (M = 0.58, SD = 0.26), t(43) = -0.54, $p > .05$, $B_{01} = 5.33$, d = 0.06. Therefore, it seems like this explanation is invalid. Indeed, Experiments 1 and 2 found a larger effect using a between-subjects design, which suggests that the absence of this effect using a within-subject design stems from carry-over effects. In the next chapter, I focus on another variable that may influence lookup behavior: memory self-efficacy.

CHAPTER 3: INDIVIDUAL DIFFERENCES IN MEMORY SELF-EFFICACY

As mentioned earlier, the retrieval aversion hypothesis states that older adults rely less on retrieval than young adults, even when their memory is sufficient, and retrieval is possible. One potential reason for why older adults avoid retrieval is that they have lower confidence in their memory ability (e.g., see Lineweaver & Hertzog, 1998). There is some evidence that global confidence in memory ability is correlated with retrieval use for older adults (Touron, Swaim, & Hertzog, 2007, see also Hertzog & Touron, 2011). There is also evidence that confidence in one's ability to use a retrieval strategy during learning is correlated with using a retrieval-based learning strategy, even when the information was well-learned (Touron & Hertzog, 2004a). Older adults have been found to benefit from prior task success, suggesting that experience mitigates any negative expectations that they have about their memory (Geraci & Miller, 2013).

There has been much research that has examined the relationship between confidence and memory, and what factors influence this relationship (e.g., Dunlosky & Nelson, 1992; Metcalfe & Finn, 2008; Nelson & Dunlosky, 1991; Rhodes & Castel, 2008). For example, Metcalfe and Finn (2008) looked at this by influencing whether an item was successfully retrieved on a prior test, to see if this would influence judgments of learning (JOLs) and subsequent study choice. Specifically, participants experienced two study-test cycles. Some of the items were repeated in the first study list and were thus more likely to be recalled during the first test compared to items that were only presented once. In a second study list, these items were only presented once, and the other items were repeated. During the second study list, participants gave higher JOLs to items that were repeated in the first study list, and chose to restudy these items less often, compared to items that were not repeated in the first study list. Critically, these two sets of items were equally likely to be recalled after the second study list. Thus, during the second study list, participants gave *unduly* high JOLs to a subset of items and were less likely to choose to restudy these items.

Similarly, researchers have found that overconfident learners appear to also be poorer performers (Dunlosky & Rawson, 2012), perhaps because of poor study habits or premature termination of study. The idea is that ones' study strategies will be based on their subjective confidence, and if that confidence is perfectly calibrated with their level of learning, then study will continue until it is learned. If one is overconfident, they will prematurely discontinue study. In contrast, if one is underconfident, they will simply discontinue study late—a choice that has few negative consequences. Therefore, although perfect calibration is a sort of "gold standard", this reasoning implies that overconfidence negatively affects learning, whereas underconfidence may simply reduce efficiency. This reasoning thus implies that underconfidence is the lesser of two evils. However, this conclusion is based on studies in which participants choose between restudying an item or not. When the scope of study strategies is limited to restudying or not restudying, this logic makes sense. In contrast, when learners can choose between restudy and test, it would seem that underconfidence would decrease the choice to self-test, which would hurt memory. Conversely, overconfidence would increase the choice to self-test, which would improve memory. Experiment 5 sought to test this hypothesis by giving false feedback about performance on a previous memory test, in an attempt to manipulate their memory self-efficacy.

Experiment 5: Effect of feedback on retrieval

Method

Subjects. A total of 209 introductory psychology students from the University of Illinois at Urbana–Champaign participated in this experiment for partial course credit. However, 7 did not return the next day for the memory test, yielding a total of 202 participants.

Materials. Paired associates were constructed for the study list and the lookup task. For the study list, 120 words were chosen. Each word was randomly assigned to be a cue or a target, and each target was randomly paired with a cue. This resulted in a list of 60 paired associates, which constituted the study list for all of the participants. The lookup task used the same paired associates as in Experiment 1 and 2.

Design. A between-subjects design was used. The type of feedback that was given after the first memory test was manipulated in order to affect one's memory self-efficacy. Participants were randomly assigned to receive either negative feedback, no feedback, or positive feedback.

Procedure. Participants first experienced a study-test cycle, followed by false feedback, and then completed the lookup task as described previously, but with no manipulation of the interface. That portion of the experiment corresponds exactly to the cheap condition in Experiments 1 and 2, and will not be described here again.

Pre-test. Participants were first shown a list of 60 paired associates at a rate of 3 seconds per word, with an interstimulus interval of 0.5 seconds. The presentation order of the paired associates was random across participants. This was followed by a self-paced cued recall test. Although performance was computed, this study-test cycle was administered only so that there would be a basis upon which to give them false feedback about their memory. The task was intentionally made to be difficult, so that people would not perform near ceiling. The rationale for this was that if a participant performed exceptionally well, but was told that they had performed poorly, they would be less likely to believe that feedback. Upon completion of the test, participants were either given negative feedback, no feedback, or positive feedback about their performance on the memory test. This was randomly pre-determined, and therefore was not related to performance. Participants in the negative feedback condition were told that they had scored “below average”, and that they “remembered less than 77% of people who have taken this test.” Participants in the positive feedback condition were told that they scored “above average”, and that they had “remembered more than 77% of people who have taken this test.” In order to make this feedback more salient and believable, those who were given feedback were shown a histogram with the “Number of people” on the y-axis. In the negative feedback condition, 23% of the histogram was shaded red, and 77% was shaded blue, with “23rd percentile” marked below the 23rd percentile boundary. In the positive feedback condition, 77% of the histogram was shaded red, and 23% was shaded blue, with “77th percentile” marked below the 77th percentile boundary. This was followed by the lookup task and was identical to the cheap condition in Experiment 1. In the no-feedback condition, participants began the lookup task shortly after completing the pretest.

Results

Pretest memory. Although we were not interested in memory before the feedback, these data were analyzed as a check for random assignment. A one-way (Feedback: Negative Feedback, No Feedback, Positive Feedback) between-subjects analysis of variance (ANOVA) revealed no differences between the conditions, $F(2, 199) = 0.14$, $p = .87$, $B_{01} = 17.22$. Thus, random assignment appeared to be successful.

Lookup behavior. Lookup behavior is shown in Figure 11. A 5 (Block) x 3 (Feedback) Mixed ANOVA was conducted on the lookup behavior. A main effect of Block was observed, $F(4, 796) = 565.94$, $p < .05$, indicating that lookup behavior decreased across Block. The effect of Feedback was not significant, $F(2, 199) = 0.28$, $p = .75$, nor was the Block x Feedback interaction, $F(4, 796) = 1.36$, $p = .21$. Collapsed across Block, lookup behavior did not differ between the Negative Feedback and No Feedback condition, $t(133.32) = -0.59$, $p > .55$, B_{01} =

4.68, d = 0.10, between the No Feedback and Positive Feedback condition, t(117.63) = -0.01, p > .05, B_{01} = 5.29, d = 0.00, or between the Negative and Positive Feedback condition, t(132.99) = -0.71, p > .05, B_{01} = 4.32, d = 0.12.

Memory. A one-way (Feedback) between-subjects ANOVA revealed no differences in memory for the paired associates in the lookup task, F(2, 199) = 0.09, p = .91. Memory did not differ between the Negative Feedback and No Feedback conditions, t(136.80) = -0.11, p > .91, B_{01} = 5.48, d = 0.02, between the No Feedback and Positive Feedback conditions, t(124.73) = -0.31, p > .05, B_{01} = 5.07, d = 0.05, or between the Negative Feedback and No Feedback conditions, t(133.99) = -0.43, p < .05, B_{01} = 4.99, d = 0.07. These data are shown in Figure 12. In all cases, the Bayes Factors indicate evidence that support the null hypothesis.

Discussion

The type of feedback given to a participant did not affect lookup behavior or memory. One potential problem with this experiment is that performance on the first memory test was quite low. Therefore, participants that were told that they remembered more than most students may not have believed the feedback. Indeed, 84% of the participants who received negative feedback believed that the feedback was accurate, whereas only 50% of the participants who received positive feedback believed that the feedback was accurate. Conversely, because the conditions of the study phase rendered learning exceptionally difficult, those who were told that they remembered less than most students may not have cared, given how difficult the task was. In other words, these participants may have reasoned that characteristics of the study phase were simply inadequate for learning. For an exploratory analysis, participants who reported that they believed the Negative or Positive feedback were compared separately. Collapsed across Block, those who believed the negative feedback (M = 3.15, SD = 0.94) did not engage in more lookups than those who believed the positive feedback (M = 3.17, SD = 0.71), t(76.63) = -0.15, p > .05, B_{01} = 4.33, d = 0.03. Although the same participants in the negative feedback condition (M = 0.51, SD = 0.25) remembered less than those in the positive feedback condition (M = 0.56, SD = 0.21), this difference was not significant, t(71.05) = -1.06, p > .05, B_{01} = 2.69, d = 0.22. Although it is possible that memory self-efficacy does not influence lookup behavior, it is also possible that our manipulation was not effective. In the next section, rather than attempting to manipulate memory self-efficacy, I report results from correlational analyses to examine this question using data from Experiment 3 and 4.

Correlational analysis: Memory self-efficacy and retrieval

As mentioned in Chapter 2, participants were given a questionnaire after finishing the memory test in Experiment 3 and 4. This primary purpose of this questionnaire was to measure memory self-efficacy (MSE), using the Capacity subscale of the Metamemory in Adulthood (MIA) questionnaire (Dixon, Hultsch, & Hertzog, 1988). Although there are many scales of MSE, this questionnaire was chosen because of its high reliability (Cronbach's alpha = 0.80–0.86, see Dixon et al., 1988), and because it does a good job at predicting memory (average r = 0.15, see Beaudoin & Desrichard, 2011). Participants were also asked questions from the Achievement and Anxiety scale, though there were no specific hypotheses regarding how Achievement or Anxiety would be related to lookup behavior and memory. Due to experimenter error, one of 16 questions from the Achievement subscale was omitted.

Procedure. In Experiments 3 and 4, after participants finished the cued recall test, they filled out the questionnaire, which included instructions (see Dixon et al., 1988). For each question, participants were to check one of five boxes. The response options were "Agree strongly", "Agree", "Undecided", "Disagree", "Disagree strongly". Each question on the

Capacity subscale asked about the participant's memory in specific situations. Each question on the Anxiety subscale asked questions that were more about how the participant feels about their memory or reacts to memory failures. Each question on the Achievement subscale asked questions that were more about their beliefs on the value of a having a good memory, embarrassment about memory failures, and behavior that the participant does to improve their memory. All of the questions from the Capacity subscale were asked first, followed by the Anxiety subscale, and then the Achievement subscale. For all of the analyses below, the responses were converted to 1-5 scale, and items that needed to be reverse scored were re-coded so that a high response reflected more capacity, anxiety, or achievement. For each participant, the responses for each question within a given subscale were averaged to create a composite score.

Results

The correlations for Experiment 3 are presented in Table 2. As you can see, Capacity was positively correlated with cued recall, and negatively related to lookup behavior. Not surprisingly, more lookup behavior was strongly associated with poorer performance on the cued recall test in both conditions. No other correlations were significant. A mediation analysis was conducted to examine whether MSE mediated any of the relationship between lookup behavior and memory. This was done by fitting the following equations to the data:

(1) $Y_j = \gamma_0 + \tau X_j + \varepsilon_{1j}$

(2) $Z_j = \gamma_0 + \alpha X_j + \varepsilon_{2j}$

(3) $Y_j = \gamma_0 + \tau' X_j + \beta Z_j + \varepsilon_{3j}$

where γ_0 refers to the intercept of given equation, *j* denotes the participant, *X* denotes the average number of times (from 1 to 5) that a word was looked up, *Z* denotes the composite score on the

Capacity subscale (MSE), *Y* denotes performance on the cued recall test (from 0 to 1). The total effect is denoted by τ, and the direct effect is denoted by τ'. The total effect represents a combination of the direct and indirect effects. Mediation occurs to the extent that the total effect is reduced (in magnitude), which is evident in the size of the direct effect, as well as magnitude of the indirect effect. Table 3 shows the parameter estimates derived from the 3 equations above. To test the hypothesis that MSE mediates the relationship between lookup behavior and memory, the confidence interval for the indirect effect ($\alpha\beta$) was computed.[2] This interval overlapped with zero, [-0.02, 0.00][3], indicating that MSE did not significantly mediate the relation between lookup behavior and memory ($\alpha\beta = -0.01$). Critically, though, there was a direct relationship between lookup behavior and memory after controlling for MSE. This result confirms the experimental result reported in Experiments 1 and 2: a greater reliance on looking up information reduced memory.

A second mediation analysis was conducted to examine whether lookup behavior mediated any of the relationship between MSE and memory. This analysis was motivated by the exceptionally large correlation between lookup behavior and memory. It was hypothesized that the observed correlation was larger than what has typically been reported in the literature partly because of the fact that participants were given more control over the stimuli, allowing MSE to influence lookup behavior, and by extension, consequent memory. This analysis was done using

---

[2] The confidence interval was computed using the `RMediation` package in R (see Tofighi & MacKinnon, 2011). This approach constructs the distribution of the product of two random variables (i.e. α and β), based on the estimates of α and β, their standard errors, and the correlation between the two estimates. The value of $\boldsymbol{\rho}_{\alpha\beta}$, within the range of [-0.99, 0.99], rarely changed the conclusion of the analysis, but exceptions are noted. For simplicity, the parameters reported were estimated using the $\boldsymbol{\rho}_{\alpha\beta} = 0$.

[3] This result did not overlap with zero when $\boldsymbol{\rho}_{\alpha\beta} < -0.89$.

the same equations above, and simply using MSE as *X*, and lookup behavior as *Z* (see Table 4). To test the hypothesis that lookup behavior mediates the relationship between MSE and memory, the confidence interval for the indirect effect (αβ) was computed, which overlapped with zero, [-0.02, 0.34], indicating that lookup behavior did not significantly affect the relationship between MSE and memory (αβ = 0.16).[4]

We repeated these analyses using the data from the within-subjects design of Experiment 4. Within-subject designs are not ideal for these analyses; nonetheless, all of the major effects reported above replicated, with the exception that MSE was no longer correlated with lookup behavior or cued recall. Table 5 shows the correlations between each subscale with lookup behavior, as well as with cued recall for Experiment 4. None of these correlations were significant. Not surprisingly, lookup behavior was strongly (negatively) associated with memory overall, $r(91) = -0.87$, $p < .001$. In addition, Capacity was negatively related to Anxiety, $r(91) = -0.40$, $p < .001$, and Anxiety was positively related to Achievement, $r(91) = 0.28$, $p < .01$. No other correlations between subscales were significant. A mediation analysis was conducted to examine whether MSE mediated any of the relationship between lookup behavior and memory. To test the indirect effect, a confidence interval was computed, which overlapped with zero, [-0.00, 0.00], indicating that MSE did not mediate any of the relationship between lookup behavior and memory (αβ = 0.00). These parameters are displayed in Table 6. A second mediation analysis was conducted to examine whether lookup behavior mediated any of the relationship between MSE and memory. To test the indirect effect, a confidence interval was computed, which overlapped with zero, [-0.10, 0.02], indicating that lookup behavior did not mediate any of

[4] This result did not overlap with zero when $\boldsymbol{\rho}_{\alpha\beta} < -0.49$.

the relation between MSE and memory ($\alpha\beta$ = -0.04).[5] These parameters are displayed in Table 8.

Discussion

In Experiment 3, there was a strong relationship between memory self-efficacy (MSE) and cued recall, as well as between MSE and lookup behavior. It is also worth mentioning that the observed association between MSE and memory was much larger than what is typically observed in studies that examine this relationship. More specifically, the relationship between MSE and memory tends to be around 0.15, and rarely ever gets above 0.2 (see Beaudoin & Desrichard, 2011). In contrast, Experiment 3 found a correlation of 0.29. This result suggests that while MSE is weakly associated with memory, the association becomes much larger when one has more control over their learning. That is, when one can decide whether to look information up or not, their MSE will become more influential in determining their later memory for that information. This interpretation is underscored by the fact that more lookup behavior was associated with a lower MSE composite, as well as poorer performance on the cued recall test. These associations further suggest that lower MSE results in more lookup behavior, which in turn, hurts memory. Although lookup behavior did not significantly mediate the relationship between MSE and memory, the indirect effect was numerically present despite a relatively small sample size of 66 participants.

General Discussion

This dissertation examined the factors that influence lookup behavior, and in turn, how that behavior influences performance on tasks that use memory. In Experiment 1, a time delay in accessing a target word decreased lookup behavior, which enhanced memory. In Experiment 2,

---

[5] This result did not overlap with zero when $\boldsymbol{\rho}_{\alpha\beta} > 0.81$.

rather than imposing an access delay, access cost was manipulated by increasing the display density of the grid. This also resulted in more lookup behavior in the cheap condition, and consequently, poorer memory than the costly condition. Access cost was manipulated in Experiment 3 by imposing a longer, more variable delay in the costly condition. Surprisingly, this manipulation had no effect on lookup behavior. Although there was no statistical difference in memory between the conditions, the costly condition did recall numerically more items than the cheap condition. Experiment 4 was a within-subjects version of Experiment 1, which also resulted in no differences between access cost conditions in lookup behavior, nor in cued recall. Experiment 5 examined whether memory self-efficacy (MSE) could be manipulated to influence lookup behavior by giving false feedback on a prior memory test. Lookup behavior did not differ between conditions, nor did cued recall. Finally, data from Experiments 3 and 4 were used to examine whether MSE was related to lookup behavior and/or cued recall. Using the data from Experiment 3, a strong correlation between MSE and cued recall was observed, which appeared to be mediated by lookup behavior. The data from both Experiments revealed a strongly negative correlation between lookup behavior and memory, which remained even after attempting to partial out any indirect effect of MSE.

Taken together, these experiments demonstrate that lookup behavior hurts memory, but that such behavior is difficult to discourage. Indeed, when lookup behavior was successfully discouraged in Experiments 1 and 2, memory was enhanced. Although there was no difference in lookup behavior for Experiment 3, it seems likely that the predictability of the stimulus onset interfered with encoding, which would make it more difficult to rely on ones' memory on later blocks. When collapsed across Experiments 1, 2, and 3, lookup behavior was strongly associated with cued recall in both the cheap, $r(169) = -0.91$, $p < .05$, and costly conditions, $r(172) = -0.83$,

$p < .05$. These associations were larger when the correlations were done using item as the level of analysis, $r(18) = -0.95$, $p < .05$, and $r(18) = -0.96$, $p < .05$, for the cheap and costly conditions respectively. Of course, these correlations may simply reflect subject- and item-selection effects, respectively.

However, the within-subjects design of Experiment 4, provides an opportunity to examine the relationship between lookup behavior and memory in a way that does not allow for such selection effects. This was done by correlating the difference in lookup behavior between the cheap and costly condition with the difference in cued recall between the same two conditions. Similar to the correlations in the between-subjects designs, more lookup behavior was strongly associated with poorer memory, $r(91) = -0.59$, $p < .05$. Importantly, because these measures are based on the same participant, this rules out the explanation that participants who looked up words more often also showed poorer memory. Similarly, because the items were randomly assigned to conditions, this rules out the explanation that items that were looked up more often were less likely to be remembered. Figure 13 depicts these differences in memory as a function of these differences in lookup behavior. This essentially shows that a mnemonic advantage in a given condition was often determined by the condition with less lookup behavior. However, lookup behavior was greater in the cheap condition than in the costly condition for about 38% of the participants, which underscores the difficulty of manipulating lookup behavior in a within-subjects design.

The data from Experiments 3 and 4 provided additional evidence that looking up information hurts memory for that information. In both Experiments, more lookup behavior predicted worse memory, even after controlling for MSE. Indeed, in Experiment 3, even though MSE was significantly related to memory ($r = 0.29$), and marginally related to lookup behavior (r

= -0.22), the total and direct effect between lookup behavior and memory were almost identical (r = -0.91 and -0.89, respectively).

It is worth mentioning that in Experiment 3, MSE was strongly associated with better memory and less lookup behavior. These results are especially noteworthy as the observed correlation in the current study is much larger than the average correlation between MSE and memory that has been observed in previous research (Beaudoin & Desrichard, 2011). Again, given that lower MSE was associated with more lookup behavior, and that more lookup behavior resulted in poorer cued recall, this suggests that MSE has a larger effect on memory when participants have more control over the information in their environment. Although lookup behavior did not significantly mediate the relationship between MSE and memory, the data were consistent with this interpretation.

As mentioned above, these experiments indicate the difficulty of manipulating lookup behavior. Whereas Experiments 1 and 2 successfully manipulated lookup behavior, access cost had no effect on lookup behavior in Experiments 3 and 4. This suggests that access cost does not work when a) the magnitude of the cost is not predictable, or b) when the cost is manipulated within subjects. It is also worth mentioning that although lookup behavior was less in the cheap conditions of both Experiments 1 and 2, the effect size was slightly smaller in Experiment 2, when the access cost was perceptual. Although this difference in effect sizes may be driven by sampling error, this would suggest that an access cost is most effective when the size of the cost is out of ones' control. Indeed, in terms of time, the access cost of the costly condition in Experiment 2 was much greater than in Experiment 1, which suggests that this extra time did not matter as much when the participants could exert some control over the size of this cost. Alternatively, it could be that a perceptual cost is not as costly as a time access cost, because of the relative ease of an external search, compared to an internal search (cf. Touron & Hertzog, 2004b). The failure to find an effect in Experiment 4 suggests that this effect is only observed in between-subjects designs, in which there are no carry-over effects (cf. Patrick et al., 2015). Finally, coupled with the observed relationship between MSE and memory found in Experiment 3, the failure to find any effect of false feedback in Experiment 5 suggests that it is difficult to manipulate ones' MSE. This could be due to the difficulty of finding a task in which subsequent false feedback is believable, or simply be because MSE is inherently stable, and difficult to influence.

As technology makes it easier to readily access more information, the opportunity to rely on this technology will increase. In other words, technology has substantially reduced the cost of accessing information. If the access cost is minimal, then it may be more rational and adaptive to look up the information on an external source. In contrast, relying on one's own memory involves searching their memory for some indefinite amount of time, with the possibility of retrieval failure, or retrieval of erroneous information. However, when one considers the fact that retrieval strengthens memory for the queried information, it follows that such dependence can hurt memory.

This relationship becomes even more concerning when one considers how the difficulty of a successful retrieval is positively related to its potency (Benjamin & Tullis, 2010; Gardiner, Craik, & Bleasdale, 1973). Given that difficult retrieval is often associated with longer retrieval latencies (Benjamin & Bjork, 1996; Karpicke & Roediger, 2007b; Koriat & Ma'ayan, 2005; Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982; Pyc & Rawson, 2009), it would follow that the immediate accessibility of information would be the most deleterious for weak memories. Similarly, individuals with lower MSE appear to be more vulnerable to the negative effect of easily accessible information than individuals with higher MSE. Thus, the effect of increased accessibility from external sources not only discourages retrieval but may also exacerbate already existing differences in memory between items, as well as individuals.

External aids serve as important tools that help us remember, and guide our behavior throughout the day. For example, we write down phone numbers, to-do lists, and reminders, as a way to offload some the memory burden, and help us accomplish our intellectual goals. As such, it is important to note that relying on external aids is not necessarily new. Indeed, perhaps one of the oldest external aids is other people, such as friends, family, or co-workers. However, with the introduction of the Internet and smart-phone technology comes the incredible ease of accessing such information, as well as the vast amount of knowledge that is available to access. That is, in response to some query, there may be some uncertainty about which of your friends, if any, has the information that one is searching for, as well as uncertainty about its accuracy. Further, one's friend may not be easily accessible or amenable to helping. Therefore, while the tendency to rely on others for information is not new, what has changed is that we now have the luxury to rely on a very knowledgeable entity who is almost always accessible, or our own personal "Data". The benefits of this technology are undoubtedly and obviously beneficial. However, it is also likely the opportunities created by this powerful technology will lead us to use our memory less, resulting in poorer memory for the queried information.

TABLES AND FIGURES

Table 1. Mixed effects model predicting cued recall

	Estimate	SE	*z*-value	*p*-value
Intercept	0.1353	0.2138	0.633	0.5268
Cost	0.4494	0.2011	2.235	0.0254

Table 2. Experiment 3: Correlation matrix with Capacity, Achievement, Anxiety, Lookup behavior, and Cued recall. The top matrix corresponds to the data that were collapsed across conditions. The middle matrix corresponds to the cheap data. The bottom matrix corresponds to the costly data.

Collapsed					
	Capacity	Achievement	Anxiety	Lookup behavior	Cued recall
Capacity	1.00				
Achievement	0.27*	1.00			
Anxiety	-0.24	0.14	1.00		
Lookup behavior	-0.22	-0.07	-0.04	1.00	
Cued recall	0.29*	0.14	0.04	-0.91***	1.00

Table 3. Experiment 3: Mediation analysis with MSE as a mediator of the relation between lookup behavior and memory.

Effect	Parameter	Estimate	SE	t-statistic	p
	α	-0.08	0.05	-1.781	.0796
	β	0.08	0.04	1.821	.0733
Total effect	τ	-0.28	0.02	**-17.19**	**<.0001**
Direct effect	τ'	-0.27	0.02	**-16.683**	**<.0001**
Indirect effect	$\alpha\beta$	-0.01	0.01		

Table 4. Experiment 3: Mediation analysis with lookup behavior as a mediator of the relation between MSE and memory.

Effect	Parameter	Estimate	SE	t-statistic	p
	α	-0.59	0.33	-1.781	.0796
	β	-0.27	0.02	**-16.683**	**<.0001**
Total effect	τ	0.24	0.10	**2.416**	**.0185**
Direct effect	τ′	0.08	0.04	1.821	.0733
Indirect effect	αβ	0.16	0.09		

Table 5. Experiment 4: Correlations between each subscale with lookup behavior, as well as with cued recall.

Collapsed		
	Lookup behavior	Cued recall
Capacity	0.08	-0.05
Achievement	0.02	0.03
Anxiety	-0.02	0.04

Table 6. Experiment 4: Mediation analysis with MSE as a mediator of the relation between lookup behavior and memory.

Effect	Parameter	Estimate	SE	t-statistic	p
	α	0.04	0.03	1.306	.193
	β	0.01	0.02	0.515	.607
Total effect	τ	-0.28	0.01	**-26.910**	**<.0001**
Direct effect	τ'	-0.28	0.01	**-26.828**	**<.0001**
Indirect effect	$\alpha\beta$	0.00	0.00		

Table 7. Experiment 4: Mediation analysis with lookup behavior as a mediator of the relation between MSE and memory.

Effect	Parameter	Estimate	SE	t-statistic	p
	α	0.15	0.11	1.306	.193
	β	-0.28	0.01	**-26.828**	**<.0001**
Total effect	τ	-0.03	0.04	-0.837	.403
Direct effect	τ'	0.01	0.02	0.515	.607
Indirect effect	$\alpha\beta$	-0.04	0.03		

Figure 1. An example trial of the noun-pair lookup task.

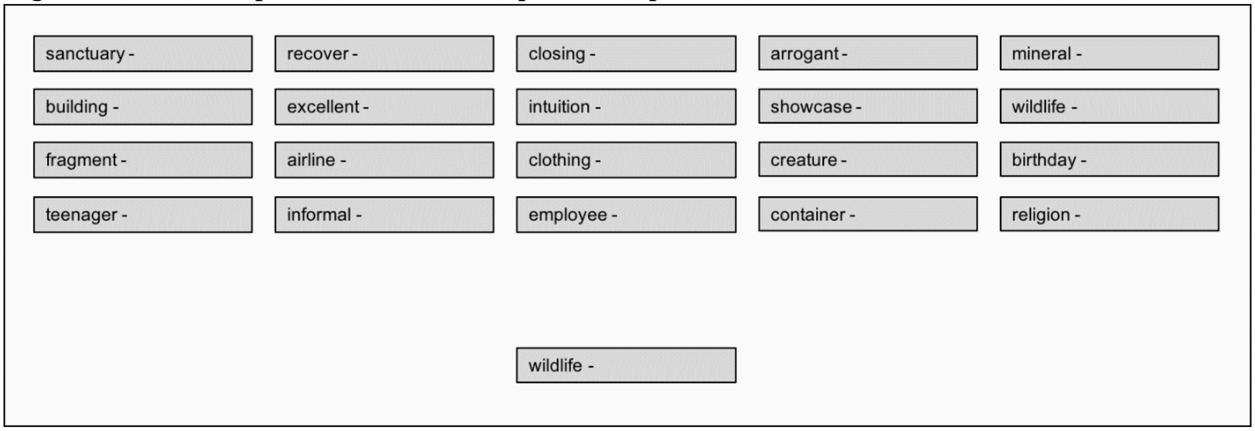


Figure 2. Experiment 1. Lookup behavior as a function of Block and Cost condition.

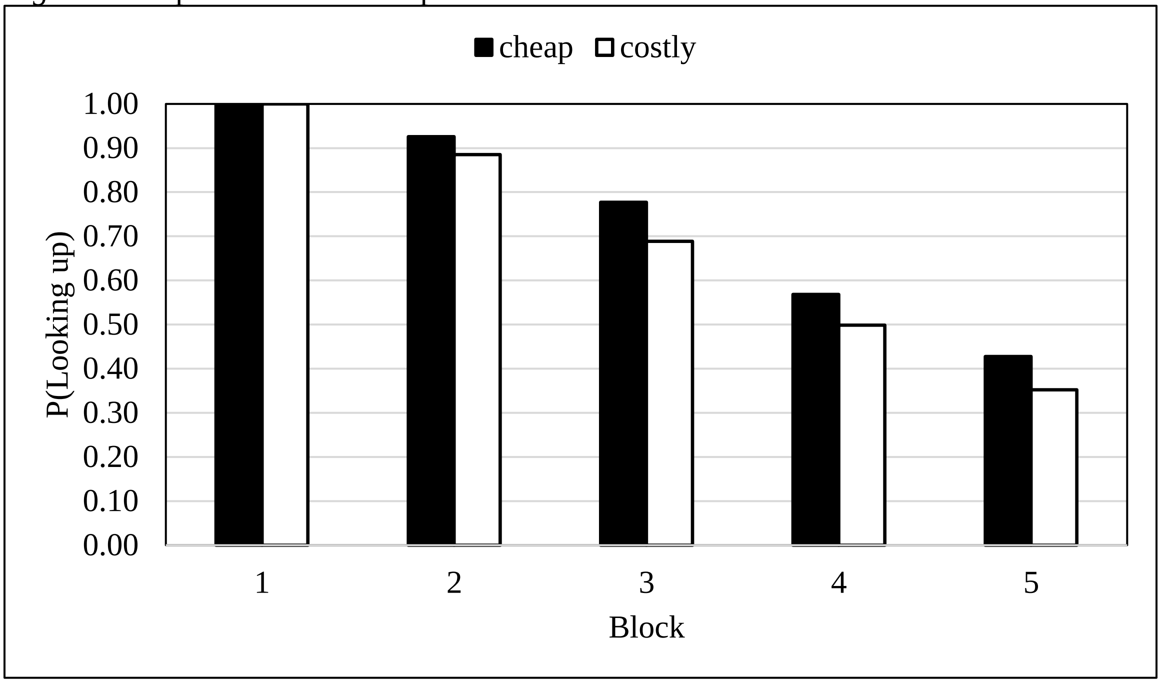


Figure 3. Experiment 1. Top panel: Memory performance as a function of Cost condition.

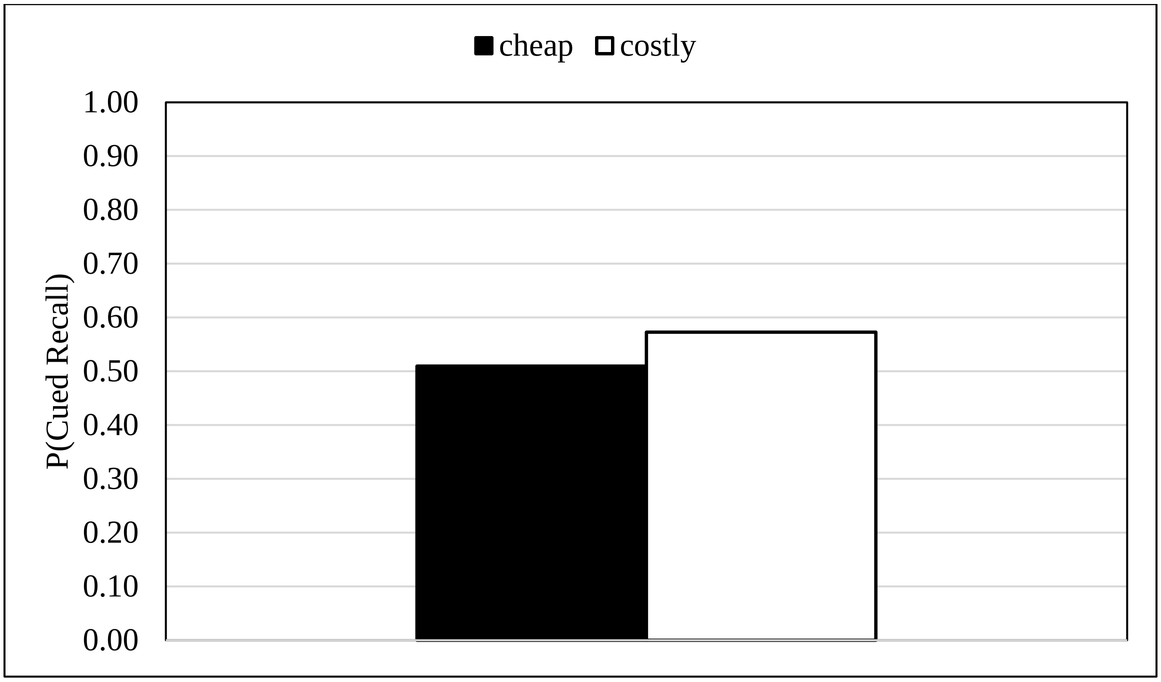


Figure 4. Experiment 2. Lookup behavior as a function of Block and Cost condition.

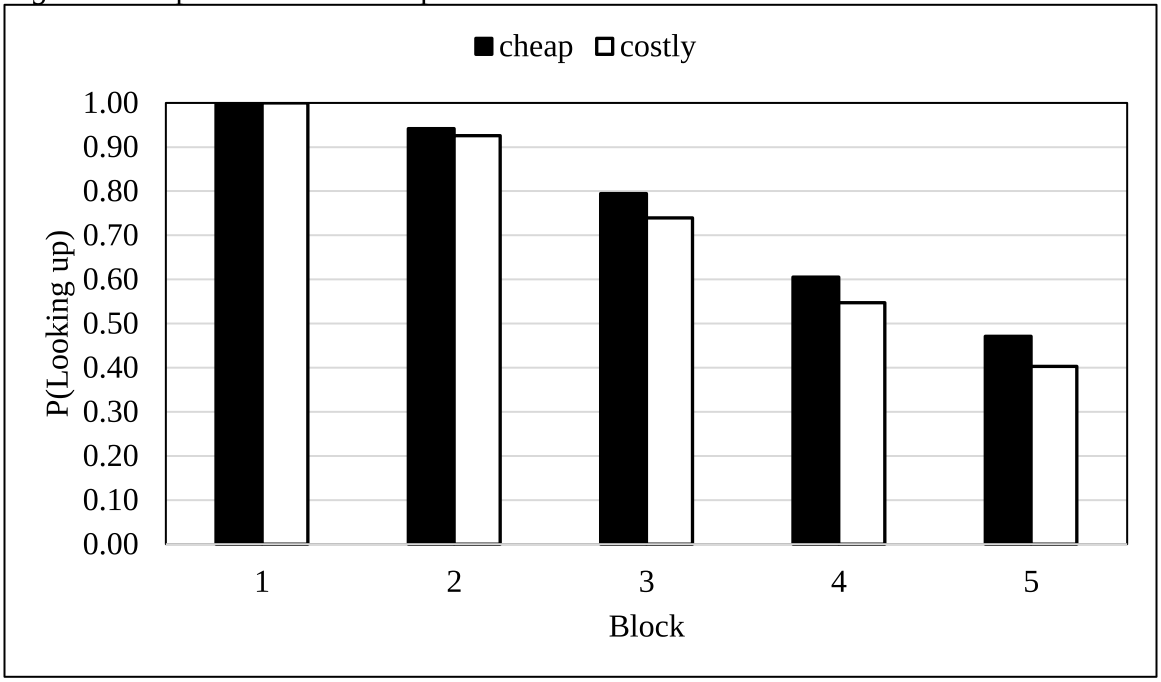


Figure 5. Experiment 2. Top panel: Memory performance as a function of Cost condition.

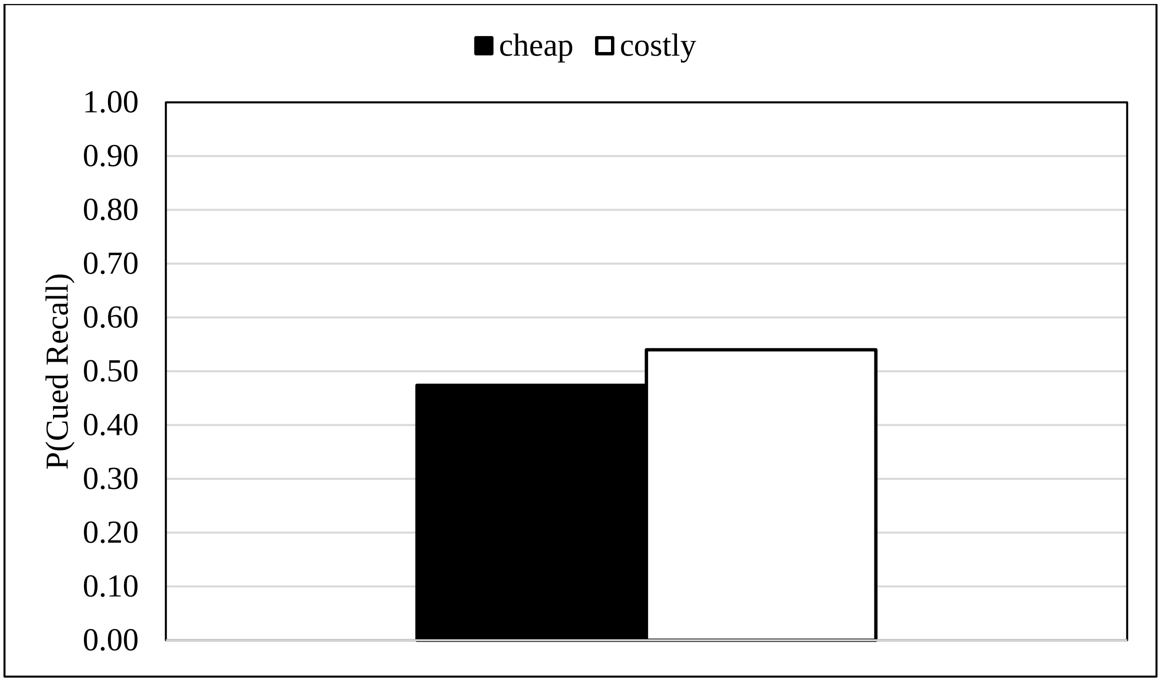


Figure 6. Experiment 3. Lookup behavior as a function of Block and Cost condition.

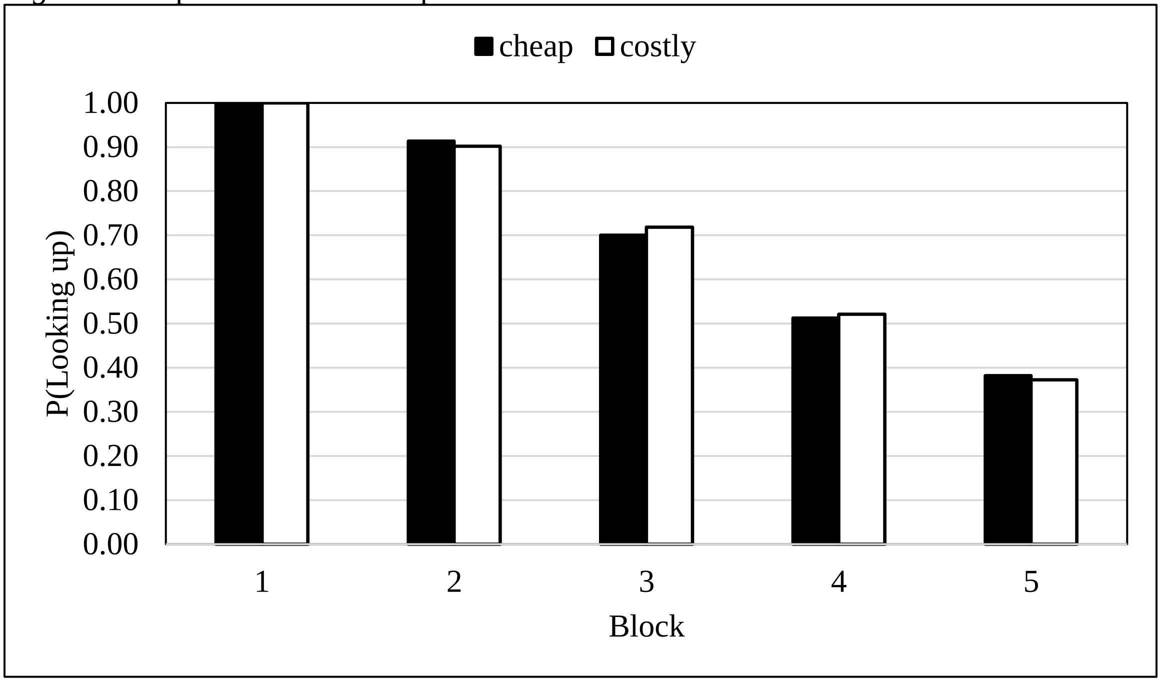


Figure 7. Experiment 3. Memory performance as a function of Cost condition.

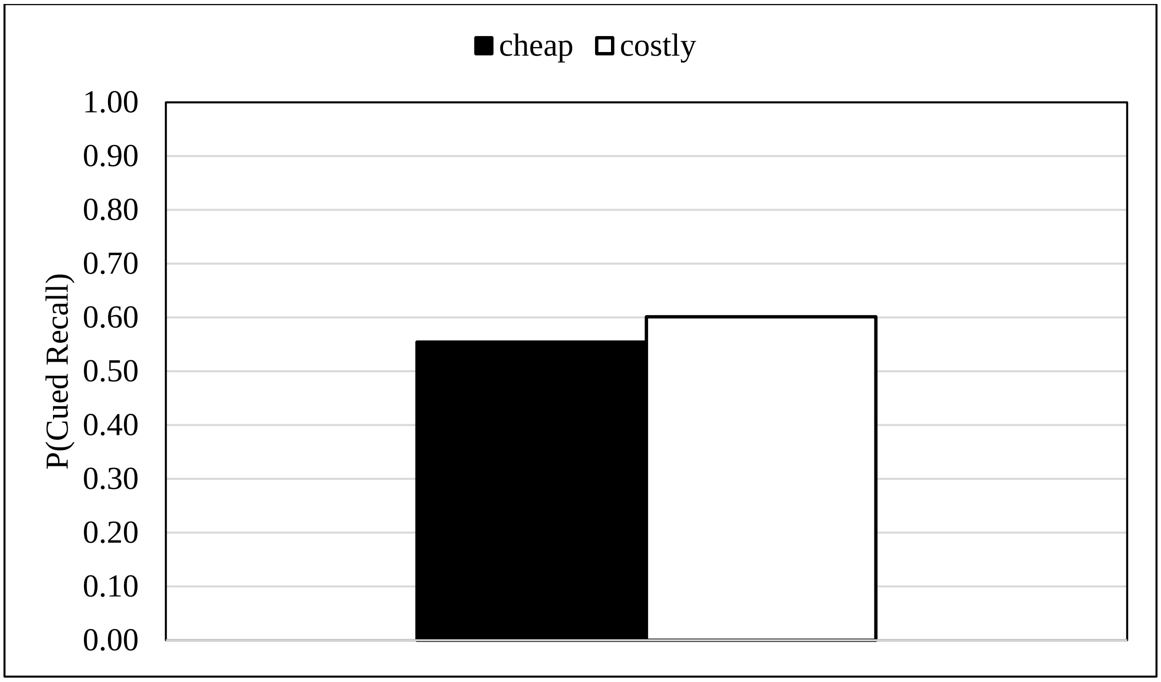


Figure 8. An example trial of the noun-pair lookup task. For Experiment 4.

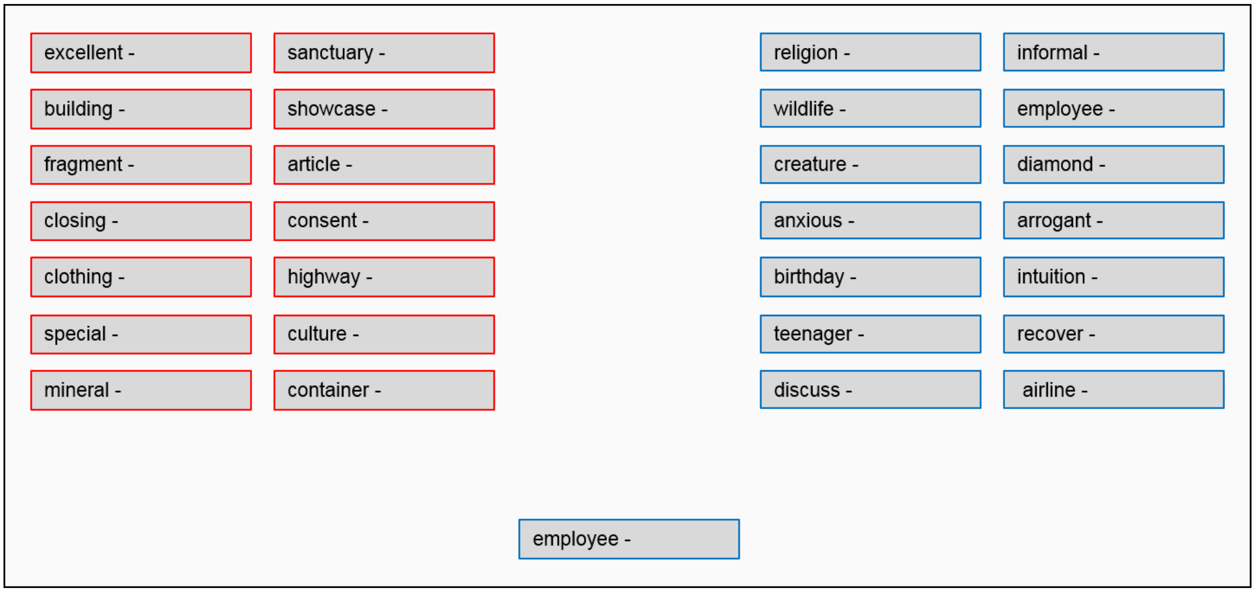


Figure 9. Experiment 4. Lookup behavior as a function of Block and Cost condition.

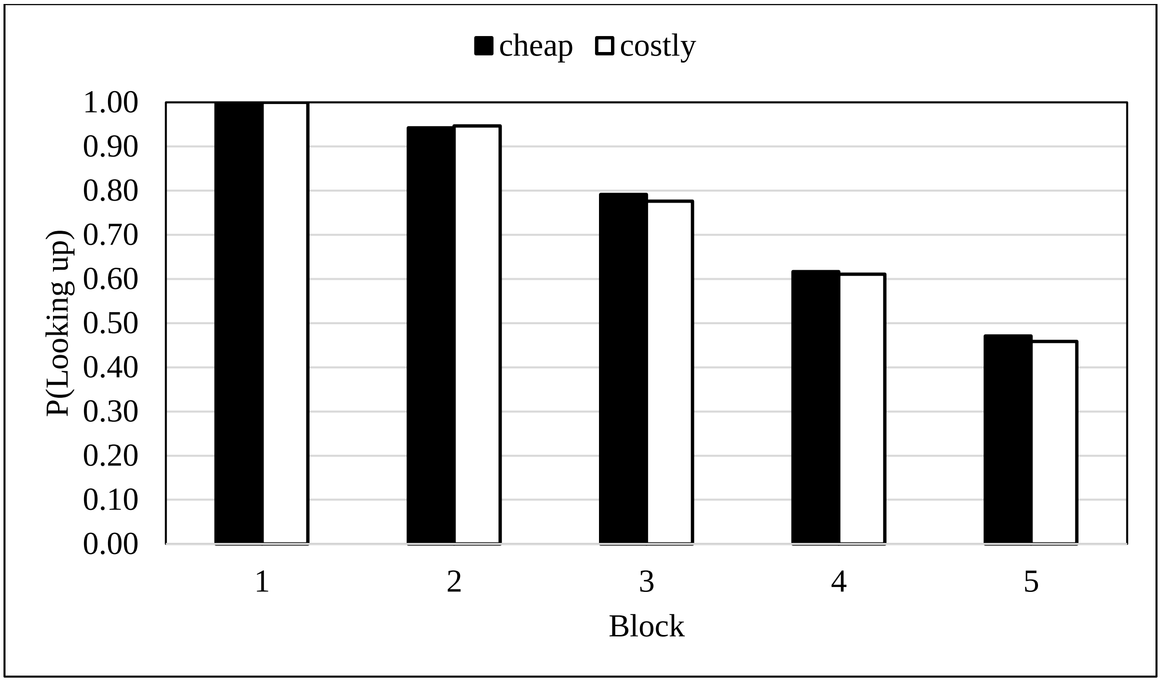


Figure 10. Experiment 4. Memory performance as a function of Cost condition.

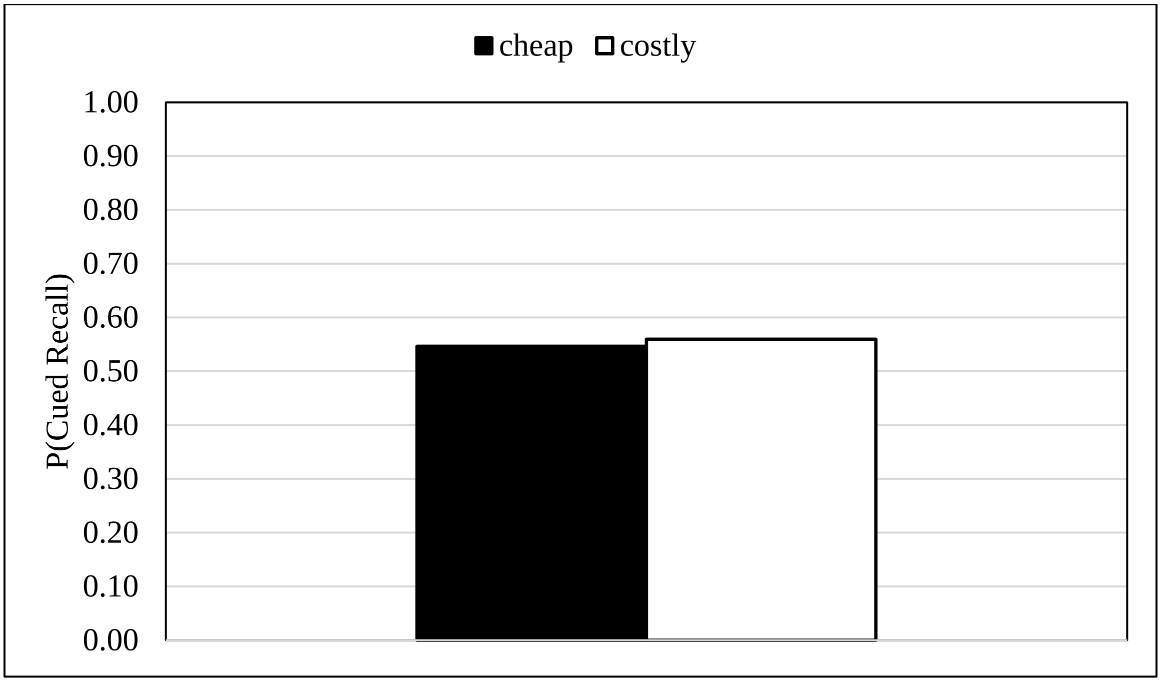


Figure 11. Experiment 5. Lookup behavior as a function of Block and Feedback condition.

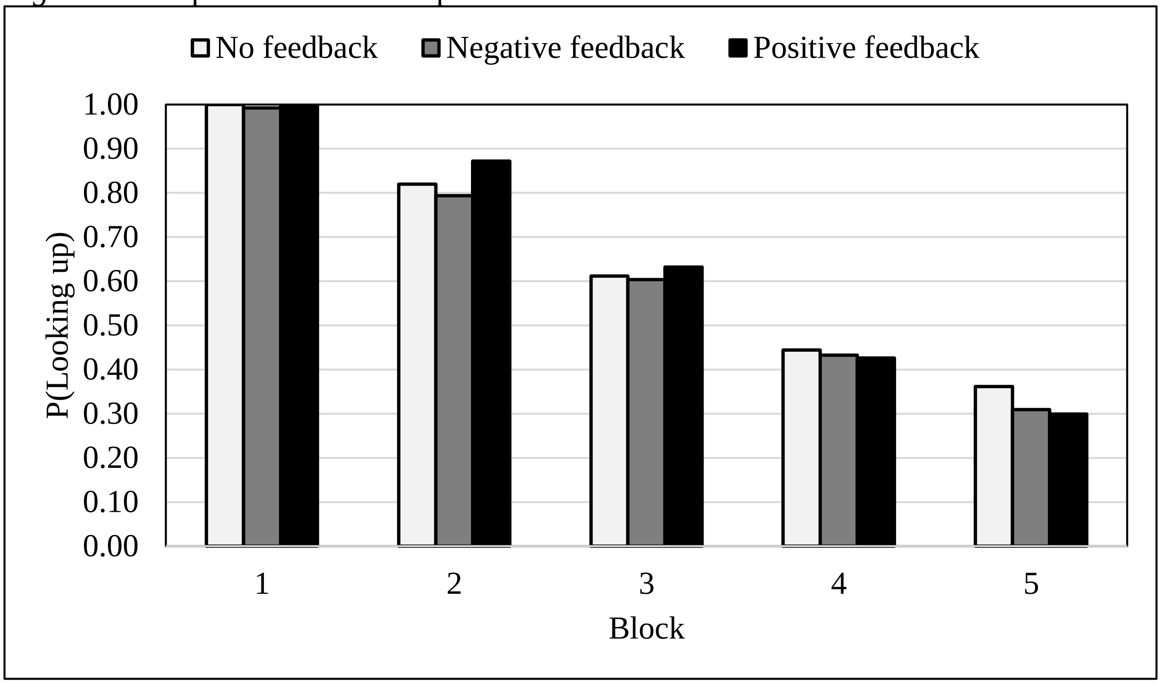


Figure 12. Experiment 5. Memory performance as a function of Block and Feedback condition.

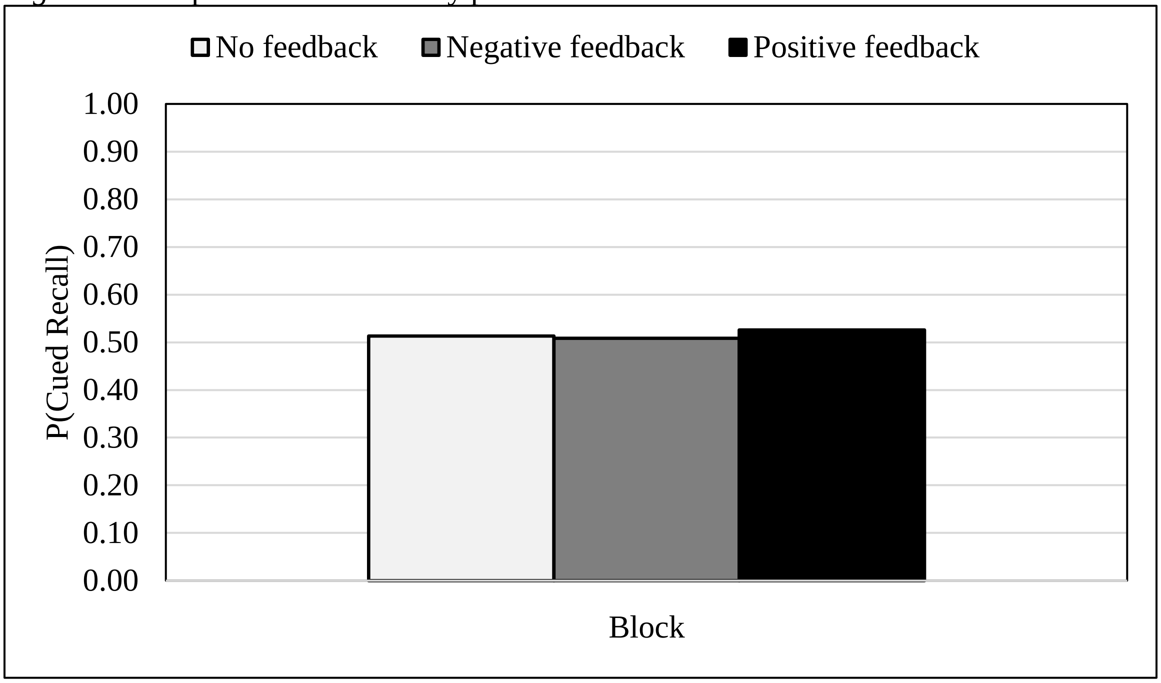
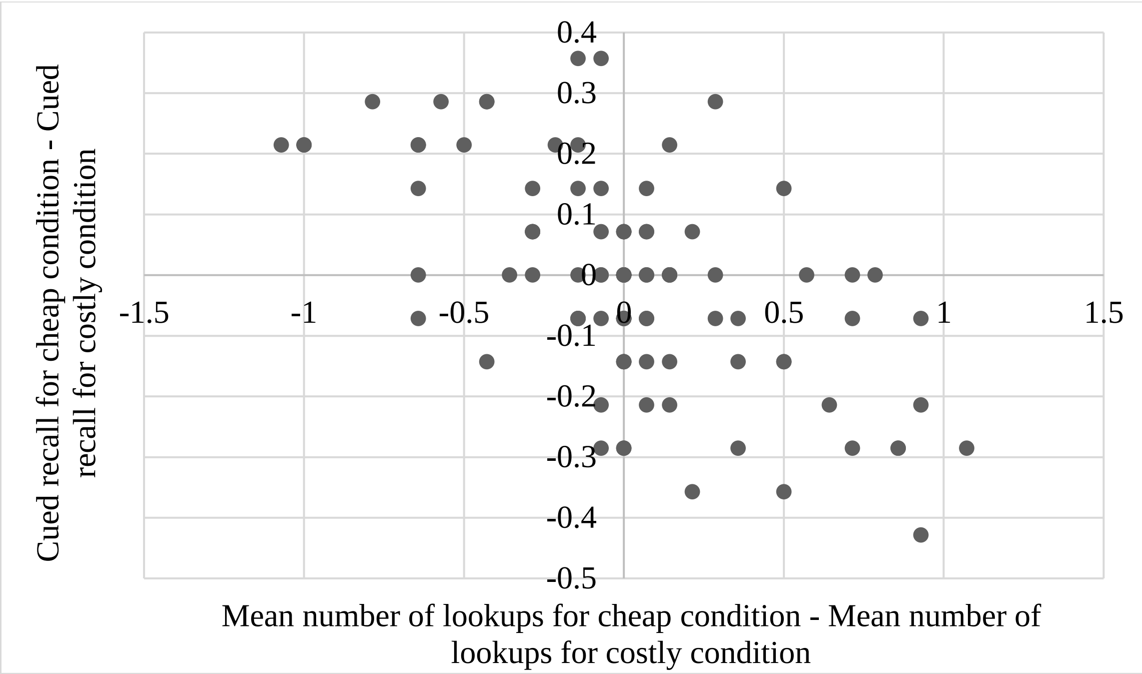


Figure 13. Experiment 4. Differences in cued recall between the cheap and costly condition as a function of differences in lookup behavior between the cheap and costly condition.



REFERENCES

Ackerman, P. L., & Woltz, D. J. (1994). Determinants of learning and performance in an associative memory/substitution task: Task constraints, individual differences, volition, and motivation. *Journal of Educational Psychology*, 86, 487-515.

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). *Applied Cognitive Psychology*, *22*, 861–876.

Ariel, R. A., & Karpicke, J. D. (2017). Improving self-regulated learning with a retrieval practice intervention. *Journal of Experimental Psychology: Applied*, in press.

Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin, 137*, 211–241.

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

Benjamin, A. S. (2005). Response speeding mediates the contribution of cue familiarity and target retrievability to metamnemonic judgments. *Psychonomic Bulletin & Review, 12*, 874–879.

Benjamin, A. S. & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. Reder (Ed.), *Implicit Memory and Metacognition.* Mahwah, NJ: Erlbaum.

Benjamin, A. S. & Pashler, H. (2015). The value of standardized testing: A perspective from cognitive psychology. *Policy Insights from the Behavioral and Brain Sciences, 2*, 13–23.

Benjamin, A. S., & Tullis, J. G. (2010). What makes distributed practice effective? *Cognitive Psychology, 61(3)*, 228-247.

Carpenter, S. K. & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, *19*, 619–636.

Carpenter, S. K. & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent attention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*, 268–276.

Carpenter, S. K. & Kelly (2012). Tests enhance retention and transfer of spatial learning. *Psychonomic Bulletin & Review, 19*, 443–448.

Carpenter, S. K. & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, *14*, 474–478.

Carpenter, S., Pashler, H., Wixted, J., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*, 438-448

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 633–642.

Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*, 553–571.

Cornell, E. H. (1980). Distributed study facilitates infants' delayed recognition memory. *Memory & Cognition*, *8*, 539–542.

Divis, K. M., & Benjamin, A. S. (2014). Retrieval speeds context fluctuation: Why semantic generation enhances later learning but hinders prior learning. *Memory & Cognition*, *42*, 1049–1062.

Dixon, R. A., Hultsch, D. F., & Hertzog, C. (1988). The Metamemory in Adulthood (MIA) questionnaire. *Psychopharmacology Bulletin, 24*, 671–688.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments-of-learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*, 373–380.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction, 22,* 271–280.

Dunlosky, J., & Rawson, K. A. (2015). Practice tests, spaced practice, and successive relearning: Tips for classroom use and for guiding students' learning. *Scholarship of Teaching and Learning in Psychology, 1,* 72-78.

Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology, 84,* 5–17.

Estes, Z., & Felker S. (2012). Confidence mediates the sex difference in mental rotation performance. *Archives of Sexual Behavior, 41*, 557–570.

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*, 632-652.

Fraundorf, S. H., Hourihan, K. L., Peters, R. A., & Benjamin, A. S. (2018). Age differences in item recognition memory sensitivity and criteria: A meta-analysis. Manuscript under review.

Fu, W.-T., & Gray, W. D. (2000). Memory versus perceptual-motor tradeoffs in a blocks world task. *Proceedings of the 24th annual meeting of the Cognitive Science Society* (pp. 310-315). Hillsdale, NJ: Erlbaum.

Gardiner, J. M., Craik, F. I. M., & Bleasdale. F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition*, *1*, 213-216.

Geraci, L., & Miller, T. M. (2013). Improving older adults' memory performance using prior task success. *Psychology and Aging, 28*, 340–345.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of*

Educational Psychology, *81*, 392–399.

Gray, W. D., & Fu, W.-T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge-in-the-head. *Cognitive Science, 28(3)*, 359–382.

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review, 113*, 461-482.

Hertzog, C., & Touron, D. R. (2011). Age differences in memory retrieval shift: Governed by feeling-of-knowing? *Psychology and Aging*, *26*, 647–660.

Hines, J., Hertzog, C., & Touron, D. R. (2012). A prelearning manipulation falsifies a pure associational deficit account of retrieval shift during skill acquisition *Aging Neuropsychology and Cognition*, *19*, 449–478.

Hertzog, C. & Hultsch, D. F. (2000). Metacognition in adulthood and old age. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 417–466). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*, 649–667.

Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1441–1451.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, *138*, 469-486.

Karpicke, J. D. & Blunt, J. R. (2011). Retrieval practices produces learning than elaborative studying with concept mapping. *Science*, *331*, 772–775.

Karpicke, J. D., Butler, A.C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory*, *17*, 471–479.

Karpicke, J. D., & Roediger, H. L. (2007a). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151–162.

Karpicke, J. D., & Roediger, H. L. (2007b). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 704-719.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966–968.

Kelemen, W. L., Winningham, R. G., & Weaver, C. A. (2007). Repeated testing sessions and scholastic aptitude in college students' metacognitive accuracy. *European Journal of Cognitive Psychology*, *19*, 689-717.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*, 219–224.

Kornell, N., & Bjork, R. A. (2008b). Optimising self-regulated study: The benefits—and costs—of dropping flashcards. *Memory*, *16*, 125–136.

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory*, *17,* 493–501.

Koriat & Ma'ayan (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language, 52,* 478-492.

Landauer, T. K. & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M.

Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.

Lineweaver, T. T., & Hertzog, C. (1998). Adults' efficacy and control beliefs regarding memory and aging: Separating general from personal beliefs. *Aging, Neuropsychology, & Cognition, 5*, 264–296.

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology*, *27*, 360–372.

Metcalfe, J., & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1084–1097.

Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive-science based program to enhance study efficacy in a high and low risk setting. *European Journal of Cognitive Psychology, 19*, 743-768.

Morgan, P. L., & Patrick, J. (2013). Paying the price works: Increasing goal access improves problem solving and mitigates the effect of interruption. *Quarterly Journal of Experimental Psychology, 66(1), 160-178.*

Morgan, P. L., Patrick, J., & Tiley, L. (2013). Improving the effectiveness of an interruption lag by inducing a memory-based strategy. *Acta Psychologica, 142(1),* 87-95.

Morgan, P. L., Patrick, J., Waldron, S. M., King, S. L., & Patrick, T. (2009). Improving memory after interruption: Exploiting soft constraints and manipulating information access cost. *Journal of Experimental Psychology: Applied, 15(4),* 291-306.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subse- quent recall: The "delayed-JOL effect." *Psychological Science, 2*, 267-270.

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 287-297.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*, 437-447.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330,* 335.

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, *140,* 283-302.

Rawson, K. A., & Touron, D. R. (2009). Age-related differences in practice effects during reading comprehension: Older adults are slower to shift from computation to retrieval. *Psychology and Aging*, *24,* 423–437.

Reder L. M. (1982). Plausibility judgments vs. fact retrieval: Alternative strategies for sentence verification. *Psychological Review*, *89*, 250–280.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology, 19(1)*, 90–138.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*, 615–625.

Roediger, Putnam, & Smith (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), *Psychology of learning and motivation: Cognition in education* (pp. 1–36). Oxford: Elsevier.

Roediger, H. L. & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210.

Roediger, H. L. & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.

Rogers, W. A., & Gilbert, D. K. (1997). Do performance strategies mediate age-related differences in associative learning? *Psychology and Aging*, *12*, 620–633.

Rogers, W. A., Hertzog, C., & Fisk, A. D. (2000). An individual differences analysis of ability and strategy influences: Age-related differences in associative learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 359–394.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*, 641–656.

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1392–1399.

Storm, B. C., Stone, S. M., & Benjamin, A. S. (2017). Using the Internet to access information inflate future use of the Internet to access other information. *Memory*, *25(6)*, 717–723.

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, *56*, 252–257.

Touron, D. R., & Hertzog, C. (2004a). Distinguishing age differences in knowledge, strategy use, and confidence during strategic skill acquisition. *Psychology and Aging*, *19*, 452–466.

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

Touron, D. R., Hertzog, C., & Frank, D. (2011). Eye movements and strategy shift in skill acquisition: Adult age differences. *The Journal of Gerontology, Series B: Psychological & Social Sciences*, *66B*, 151–159.

Touron, D. R., Hoyer, W. J., & Cerella, J. (2001). Cognitive skill acquisition and transfer in younger and older adults. *Psychology and Aging*, *16*, 555–563.

Touron, D. R., Hoyer, W. J., & Cerella, J. (2004). Age-related differences in the component processes of cognitive skill learning. *Psychology and Aging*, *19*, 565–580.

Touron, D. R., Swaim, E., & Hertzog, C. (2007). Moderation of older adults' retrieval reluctance through task instructions and monetary incentives. *The Journals of Gerontology, Series B: Psychological Sciences & Social Sciences*, *62*, 149–155.

Tullis, J. G., Fiechter, J. L., & Benjamin, A. S. (2017). The efficacy of learners' testing choices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44,* 540-552.

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.

Waldron, S. M., Patrick, J., & Duggan, G. B. (2011). The influence of goal-state access cost on planning during problem solving. *Quarterly Journal of Experimental Psychology, 64(3),* 485-503.

Waldron, Patrick, & Morgan (2007). Influencing cognitive strategy by manipulating information access costs. *The Computer Journal*, *50*, 694-702.

APPENDIX A: GENDER DIFFERENCES

This Appendix includes an exploratory analysis that examines gender differences. This analysis was originally motivated by literature suggesting that males are more confident than females (Ehrlinger & Dunning, 2003; Estes & Felker, 2012), and thus may have higher memory self-efficacy. This starting point led us to predict that males would show less lookup behavior than females. This prediction was not confirmed, though we did find some evidence that males were more sensitive to the manipulation of access cost. The following analyses were conducted on the data combined across Experiments 1, 2, and 3.

Lookup behavior. A 5 (Block) x 2 (Cost) Mixed ANOVA was conducted on the lookup data, revealing a main effect of Block, $F(4, 1372) = 858.93$, $p < .05$, indicating that participants looked up words less often as a function of Block. There was also was marginally significant main effect of Cost, $F(1, 343) = 3.77$, $p = .0529$, indicating that the costly conditions looked up words less often than the cheap conditions. There was also a marginally significant interaction, $F(4, 1372) = 2.21$, $p = .0661$, indicating slight tendency for participants in the costly conditions to rely on retrieval from memory more as a function of Block. Collapsed across Block, the cheap conditions looked up slightly more ($M = 3.71$, $SD = 0.90$) than the costly conditions ($M = 3.51$, $SD = 0.94$, though, this was only marginally significant, $t(342.81) = 1.95$, $B_{01} = 1.36$, $d = 0.21$, $p = .05192$. These data are shown in Figure A1.

A 5 (Block) x 2 (Cost) Mixed ANOVA was conducted on the lookup behavior of females, revealing a main effect of Block, $F(4, 892) = 605.21$, $p < .05$, indicating a reduction in lookup behavior across Block. There was no main effect of Cost, $F(1, 223) = 0.53$, $p > .05$, nor was there a Block x Cost interaction, $F(4, 892) = 0.37$, $p > .05$. Collapsed across Block, lookup behavior did not differ between the cheap (M = 3.63, SD = 0.90) and the costly conditions (M = 3.55, SD = 0.89), t(222.7) = 0.73, B_{01} = 5.36, d = 0.10, p > .05.

A 5 (Block) x 2 (Cost) Mixed ANOVA was conducted on lookup behavior of males, revealing a main effect of Block, F(4, 460) = 251.03, p < .05, indicating less lookup behavior as a function of Block. The was a main effect of Cost, F(1, 115) = 4.48, p < .05, indicating less lookup behavior in the costly conditions. There was also a significant Block x Cost interaction, F(4, 460) = 2.92, p < .05, indicating a greater reduction in lookup behavior as a function of Block in the costly conditions. Collapsed across Block, the cheap conditions looked up target words more often (M = 3.84, SD = 0.92) than the costly conditions (M = 3.46, SD = 1.02), t(114.14) = 2.15, B_{10} = 1.56, d = 0.40, p < .05. These data are displayed in Figure A2.

Memory. Performance on the cued recall test was higher in the costly conditions (M = 0.57, SD = 0.26) than in the cheap conditions (M = 0.50, SD = 0.29), t(337.57) = -2.06, B_{10} = 0.91, d = 0.22, p < .05. Females did not differ in cued recall between costly (M = 0.58, SD = 0.25) and cheap conditions (M = 0.54, SD = 0.28), t(221.04) = -0.90, B_{01} = 4.70, d = 0.12, p > .05. In contrast, males in the costly conditions recalled more (M = 0.55, SD = 0.28), than in the cheap conditions (M =0.42 , SD = 0.29), t(109) = -2.33, B_{10} = 2.23, d = 0.44, p < .05. These data are shown in Figure A3.

Taken together, it appears that this manipulation has a much larger effect on lookup behavior for males (d = 0.40) than for females (d = 0.10). Not surprisingly, this manipulation also had a larger effect on cued recall for males (d = 0.44) than for females (d = 0.12). Therefore, the effect that is seen when collapsed across experiments is almost entirely driven by males.

Figure 14. Experiments 1–3. Lookup behavior as a function of Block and Cost condition.

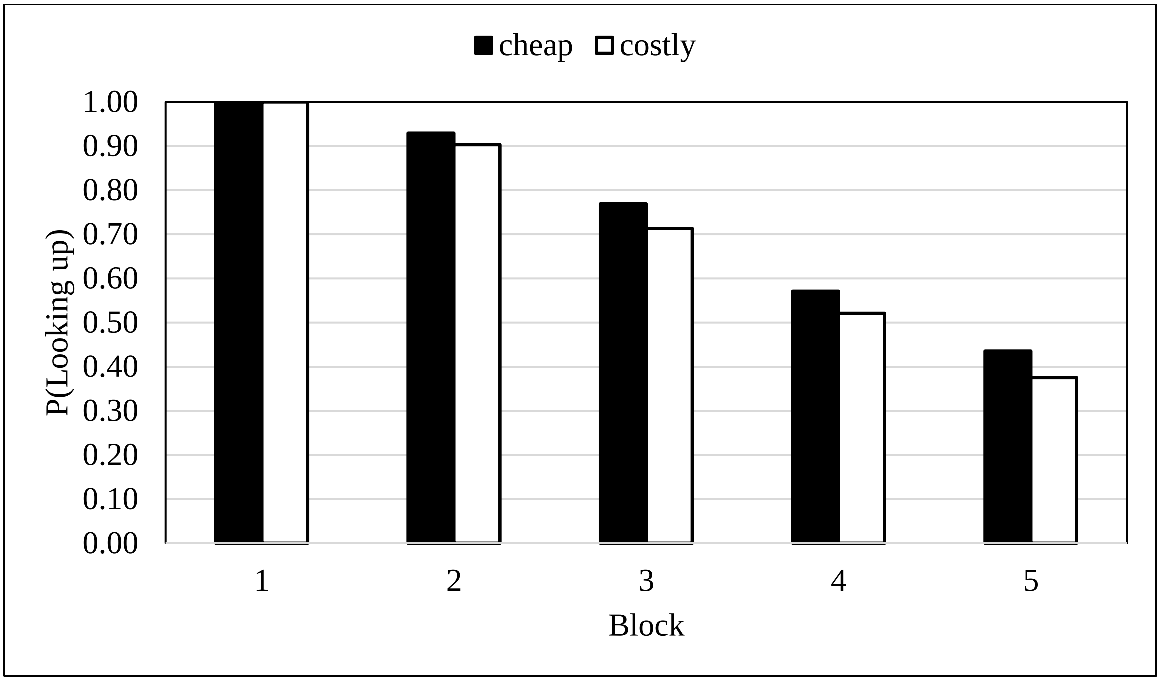


Figure 15. Top panel: Lookup behavior as a function of Block and Cost condition for females. Bottom panel: Lookup behavior as a function of Block and Cost condition for males.

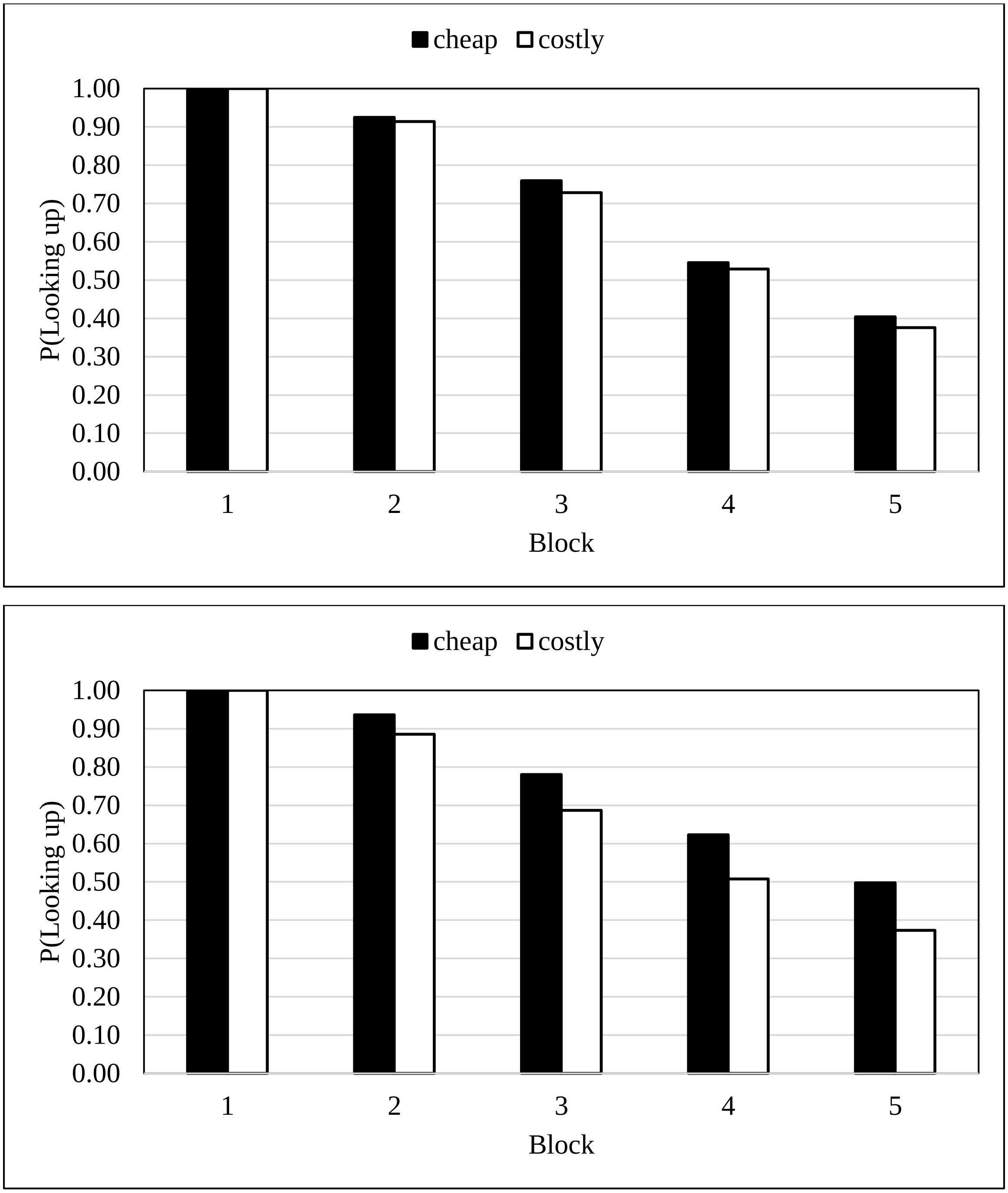


Figure 16. Top panel: Memory performance as a function of Cost condition. Bottom panel: Memory performance as a function of Cost condition and Gender.