UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE FORWARD EFFECT OF TESTING ON COMPREHENDING COMPLEX TEXTS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

THOMAS BOBBITT
Norman, Oklahoma
2018

THE FORWARD EFFECT OF TESTING ON COMPREHENDING COMPLEX TEXTS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF PSYCHOLOGY

BY

Dr. Scott Gronlund, Chair

Dr. Daniel Kimball

Dr. Michael Wenger

Dr. Robert Terry

Dr. Sepideh Stewart

Table of Contents

List of Tables

List of Figures

Abstract

Research on the forward effect of testing has indicated that taking a test over prior material can improve one's ability to memorize new material. However, no research has yet indicated whether such testing can improve one's ability to understand or comprehend new complex prose material. Theory and data from the comprehension literature, testing effect literature, and metacomprehension literature suggest that test-taking may have this forward effect on one's ability to understand complex prose material. A series of experiments are conducted that tested this possibility. In each experiment, participants (1) read a text, (2) reprocessed that first text in some way, (3) read a second text that was related to the first text, (4) took an inference test over the second text, which was used to assess their comprehension of it. It was hypothesized that answering inference questions over an initial text would better enable one to understand a subsequent, related text than would rereading the initial text (Experiments 1 and 2) or answering memory questions over the initial text (Experiment 3). It was also hypothesized that answering these inference questions would be more effective if the answer-choices were accompanied by arguments than if they were not (Experiment 4). Ultimately, only the hypothesis of Experiment 4 was supported. Ps likely must have a certain degree of motivation to learn in order for there to be a forward effect of testing on comprehension. Nonetheless, it is argued that inference questions should be used more frequently in education.

Keywords: testing effect; comprehension; metacomprehension

Forward Effect of Testing on Comprehending Complex Texts

Cognitive psychologists recently have proposed that educators dramatically increase the number of tests and quizzes given in classes, not only because test-taking is a potent learning event (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Pashler, et al., 2007; Roediger, Putnam, & Smith, 2011; Roediger & Pyc, 2012), but also because students are largely under the illusion that test-taking is not a learning event at all (R. Bjork, Dunlosky, & Kornell, 2013). However, the *kinds* of questions on these tests is important because students tend to study new class material in a way consistent with how they expect to be tested on it—the *test-expectancy effect* (McDaniel, Blischak, & Challis, 1994)—and a major predictor that students use in this regard is experience from prior tests. Thus, students might try to merely memorize information explicitly stated in lectures and textbooks if prior tests only assessed their memory for this information (Thiede, Wiley, & Griffin, 2011). This is problematic because one must have a deeper understanding of a given body of information in order to apply one's knowledge of it to novel situations, and to not have knowledge that is "inert" (W. Kintsch, 1994).

Educators presumably hope that their students will apply what they have learned outside the confines of the classroom (Momsen, Long, Wyse, & Ebert-May, 2010). Thus, it is desirable to use test questions that require learners to apply explicitly stated information to novel contexts and to deduce what is only implicit in that information. I will be referring to such questions as *inference questions*. Inference questions prompt students to try to understand subsequent texts rather than to just memorize them (Thiede et al., 2013). However, many issues remain, such as whether the act of taking inference tests has benefits other than instilling test-expectancy effects, and whether the nature of these benefits varies as a function of the type inference question used. The purpose of this dissertation is to explore these issues.

Although there has been plentiful research on the effects of testing in recent years (for a review, see Roediger et al., 2011), the vast majority of this research has been on *backward effects* of testing; research on the *forward effects* of testing has barely begun (for a review, see Pastötter & Bäuml, 2014). The backward effect of testing occurs when taking a test improves one's learning of the information that was tested. For example, Roediger and Karpicke (2006) found that participants (Ps) forgot the information contained in a science passage much more slowly if they had earlier tested themselves over that information than if they had earlier reread it. The forward effect of testing, by contrast, occurs when taking a test improves one's learning of information that is first presented *after* the test is taken. Although we will make use of the substantial research done on the backward effect of testing, the focus of this dissertation is on the relatively unexplored forward effects of testing.

In the following literature review, I will begin by describing current research on the forward effect of testing, and point out that possible forward effects of testing on comprehension (as opposed to memory) are yet to be explored. I will then briefly discussion two topics important to this consideration, namely, research on comprehension and research on the testing effect. In so doing, I will briefly describe W. Kintsch's (1994) construction-integration model as well as two different theories about the backward effect of testing. As we will see, only one of these two different testing effect theories seems to predict a forward effect of testing on comprehension. I will then argue that, nonetheless, considerations from the metacomprhension literature strongly suggest that answering *some* kinds of test questions will improve the ability to comprehend a future, related topic. There will then be a brief discussion of which kinds of test questions will be most beneficial to students in this regard.

I will ultimately form the following hypotheses. First, answering inference question over a text will result in a larger forward effect than will rereading that text. Second, answering inference question over a text will also result in less subsequent mind-wandering than will rereading that text. Third, I will hypothesize that answering inference questions will result in a larger forward effect than will answering memory questions. Lastly, I will hypothesize that answering multiple-choice inference questions *in which each answer-choice is accompanied by an argument supporting it* will result in a larger forward effect than will answering multiple-choice questions *in which each answer-choice is **not** accompanied by an argument supporting it*.

Research on Forward Effect of Testing

There have been a number of studies demonstrating that testing does indeed have a forward effect. For example, Szpunar, McDermott, and Roediger (2008) had Ps study five lists of unrelated, common nouns. Ps were better able to encode and remember the words from the fifth list if they had taken tests earlier over the first four lists than if they had merely solved math problems. Much of the research establishing the forward effect of testing does, like Szpunar et al. (2008), use simple learning material that can only be memorized and that do not admit of learning at a deeper level (e.g., Nunes & Weinstein, 2012; Tulving & Watkins, 1974; Wahlheim, 2015; Yang, Potts, & Shanks, 2017). Thus, many of the theoretical explanations of these findings will not be of much interest to this dissertation.

For example, Pastötter, Schicker, Niedernhuber, and Bäuml (2011) gave evidence from electroencephalorgram data suggesting that tests help one to maintain attention at a high level. There is also evidence that tests can cause dramatic shifts in mental contexts, meaning that P's memories for each list will be better differentiated from each other (Sahakyan & Hendricks, 2012). In other words, taking tests over the first four lists in Szpunar et al.'s (2008) study enabled Ps to encode each list as a unique episode of simple and manageable size as opposed to encoding all lists within a single, large episode. One thing to note about these kinds of explanations is how ephemeral the benefit is: taking a quiz at the end of a class today will in no way improve learning of the material presented in the next class. That next class will already be in a sufficiently different context and the main determinant of a student's attention during it will be what was transpiring right before that class. More importantly, though, is that these effects of testing do not so much *improve* one's ability to understand subsequent information as they allow one to *maintain* the ability one already had (or to prevent that ability from declining as a result of fatigue). Hence, it is so far unclear whether testing will be of substantial benefit in facilitating the understanding of more complex material.

There have so far been three studies documenting a forward effect of testing for the learning of more complex material. Szpunar, Khan, and Schacter (2013, Experiment 2) presented Ps with a 20-min lecture video over statistics, which was divided into four 5-min segments. In between each segment, Ps either (1) answered questions about information covered in the prior segment, (2) read those very same questions along with their answers, or (3) solved unrelated math problems. Once the final segment was presented, all Ps took a test over that final segment. Ps who had taken earlier tests did better on this final test than did Ps from the other groups. Jing, Szpunar, and Schacter (2016) found the same results in a similar study, in which Ps viewed a 40-min video over the sociology of health care (divided into eight 5-min segments). Importantly, both Szpunar et al. (2013) and Jing et al. (2016) found that the Ps who had to answer test questions reported less mind-wandering than did the other Ps; the authors argue that this reduction in mind-wandering is an explanation for why they observed a forward effect of testing.

Wissman, Rawson, and Pyc (2011, Experiment 3) also conducted a study in which there was a forward effect of testing in the learning of complex material. They presented Ps with a text on storing greenhouse gases, which was broken up into 3 segments (roughly 330 words per segment). In between each segment, one group attempted to recall the information presented in the prior segment, whereas the other group solved unrelated math problems. Both groups tried to recall all of the information presented in the third segment once they finished reading it. The group that had earlier attempted to recall information from the prior segments was better able to recall the information presented on the final segment than was the group that had worked on math problems.

One thing to note about the experiments just described is that they only demonstrated that tests can improve one's ability to *remember* subsequent information; none of them measured the Ps' *understanding* of or ability to make inferences from the presented information. This distinction is important because some manipulations improve memory at the *expense* of understanding (Hinze, Griffin, Wiley, 2013; Mannes & Kintsch, 1987). Nonetheless, there are reasons to expect that testing can improve the ability to understand information presented in the future.

First, there is the fact that deep understanding of a text usually requires the use of relevant background knowledge (Kintsch, 1994; McNamara & Kintsch, 1996), and the plentiful research on the backward effect of testing has established that testing can facilitate the learning of the information being tested (Roediger et al., 2011), which could become relevant background knowledge for a later text. For example, consider a fairly advanced textbook in the sciences. Often, one must have already learned the contents of the first chapter in order to fully understand the second chapter. Hence, if taking a test over the first chapter improves one's learning of it, then it might also improve one's learning of the second chapter. However, we will see that the *way* in which testing improves one's learning of the first chapter is a crucial issue that has bearing on this line of reasoning. I will argue that the metacomprehension benefits of testing—or the way in which tests facilitate a learner's ability to gauge how well they understand a text—suggest that a test over a first chapter will improve learning of the next chapter. I will elaborate on each of these two points.

Background Knowledge and the Backward Effect of Testing

The importance of background knowledge for understanding a text can be illustrated in Kintsch's (1994) Construction-Integration Model. The model distinguishes between comprehension of the *textbase* and comprehension of the *situation model*. The *textbase* refers to that which is explicitly stated in a text. For example, a text might state: "When a baby has a septal defect, the blood cannot get rid of enough carbon dioxide through the lungs." One can read this sentence and understand the textbase—that having a septal defect in some sense *causes* a problem with removing the carbon dioxide. But, without using other knowledge, it will remain unclear *why* septal defects have this effect. Hence, relating this sentence to some relevant background knowledge is essential for fully understanding it, for seeing how the two parts of the sentence relate to each other. In this case, one would need to know (1) that blood travels through the heart just before it goes to the lungs, (2) that blood travels back through the heart immediately after it has been cleansed of carbon dioxide in the lungs, and (3) that the septum is a wall in the heart that keeps separate the blood that is on its way to the lungs and blood that has already gone through the lungs. If one relates the sentence stated in the text with this background knowledge, then one can see how septal defects can causes blood to not be fully ridded of carbon dioxide, namely, that some blood that is first entering the heart will, rather than moving on to the lungs, pass instead through the disfigured septum and mix with the blood that has already gone through the lungs. This act of relating and fitting together various pieces of information, both from the text and from one's own background knowledge, is the act of constructing the *situation model* being described in the text, which is the act of understanding the text at a deep level. Background knowledge is in this way often essential for understanding complex texts. I mentioned above that taking a test over the first chapter of a textbook could improve one's ability to understand content from the next chapter if knowledge of the first chapter is necessary for understanding the second, and if a test over the first chapter improves one's learning of it. Much research on the testing effect suggests that tests will have this beneficial effect; however, the situation is in fact complicated and requires further analysis.

One can remember information that is relevant to a task and yet not notice during that task that one should remember and apply that information (Gick & Holyoak, 1980). Thus, while reading the second chapter of a textbook (or the sixth chapter), some particular content in that chapter must remind you of other relevant information that you have encoded previously if you are to think of it and integrate it with the new chapter. This reminding is more likely if that particular bit of old information is embedded within some well-constructed situation model that inter-relates a multitude of facts and concepts, for then there are many retrieval routes through which one can be reminded of that particular bit of old information (Kintsch, 1994). By contrast, if that bit of information has not been well integrated with other knowledge, and is basically just associated to episodic contexts in which it has been encoded, then one will likely have to be first reminded of that episode or have that information explicitly prompted in order to remember (W. Kintsch, 1994). Hence, if testing oneself over information only improves the ability to recall that information when explicitly prompted to, and does not facilitate its integration within a situation model, then testing may not be beneficial in helping one to understand new information at a deep level. We look next at the theoretical accounts that have been given to explain the backward effect of testing in order to examine the possibility that testing oneself over information does indeed facilitate the integration of that information within a situation model.

There are two major accounts that have been proposed to explain the testing effect: the episodic context account (ECA; Karpicke, Lehman, & Aue, 2014) and the elaborative retrieval hypothesis (ERH; Carpenter, 2009). The ECA is centered upon the concept of us having fluctuating mental contexts (Mensink & Raaijmakers, 1989; see also Estes, 1955). According to this idea, each person processes information within some particular context, which includes characteristics of mood, of their surroundings, of the time of the day, etc., *as the person experiences them*. When one is presented with information and stores it as a memory, he or she also encodes some of the features/elements of that context into the memory (encoding of the context is usually assumed to be incomplete, though). If one later tries to recall that information, he or she will often try to mentally reinstate features from the context in which the information was encoded, which will differ in some degree from the features of the new and current context. This idea makes intuitive sense if we consider a person trying to remember something that they were taught *in a specific class* and *on a specific day*. Reinstating the context in which the information was encoded does help one to remember that information due to there being an association between the two, and so one will be more likely to recall the information if one better reinstates the context in which it was encoded. However, reinstatement of the prior context is never perfect, and so some elements of the new context will remain. According to the ECA, if the memory is retrieved, then some of the elements from the new context will be encoded into that memory. As a greater variety of context elements become encoded into the memory, the better able one will be to reinstate a subset of those features at a given time in the future (see Lehman & Malmberg, 2013), and so the better able one will be to retrieve that memory. There are two reasons why testing oneself over, say, a passage is more conducive to this form of context-updating than is rereading. When you test yourself over some information, you must actually retrieve the memory of this information. This retrieval is also a necessity if one is to update that memory with elements from the new and current context. Also, the focus of one's attention differs between the two activities: because testing requires you to have to remember what happened in a prior context, your attention will be strongly focused on context elements. Thus, you will encode more of the context elements when testing than when rereading.

The logic of the ERH, like that of the ECA, is driven by the idea that remembering is a cue-driven process: the ability to remember something is largely determined by whether one is thinking of other things—"cues"—that are strongly associated with it. However, whereas the ECA focuses on context cues, the ERH focuses on semantic cues. Consider the situation in which one is asked: "Does a septal defect cause people to not get rid of as much carbon dioxide as they should?" In this scenario, we are assuming that the person in question has already been told the answer and is now trying to remember that answer. According to the ECA, prior testing could help people answer this question by facilitating their ability to reinstate a prior context in which they learned about the septum. According to the ERH, by contrast, prior testing would help by facilitating people's ability to directly remember other information related to the heart and lungs, and furthermore facilitating the ability of this other information to remind them of the answer. The reason the ERH makes these assumptions can be illustrated thus. If one is right now trying to answer this question about the septal defect, he or she will probably be trying to remember what the answer is. Because remembering is a cue-driven process, the person will be using key words from the question, such as septum, as cues to help remind them of the answer. During the period of this activity, some of the other memories associated to these cues will become activated, such as memories containing information about the course of blood flow. If one eventually remembers the answer, then not only will the association between the answer and 'septum' be strengthened, but so will the associations between the answer and the other information that became activated during the process (Carpenter, 2009; 2011). Importantly, some of this other information might be relevant background knowledge that is necessary for understanding a text at a deep level, and perhaps it is more likely to be neglected when one is reading the text than when one is trying to remember it. The ERH explanation of the testing effect is thus clearly more consistent with the idea that testing can facilitate the learning of new information than is the ECA, for it is only according to the ERH that testing seems to be conducive to the construction of situation models, which, again, help one to recognize, during later learning events, that some given information is relevant. The ECA, by contrast, seems more consistent with the idea that testing merely improves one's ability to remember some information once one already wants to remember it. Thus, it is important to determine how consistent these accounts (the ERH, in particular, for our purposes) are with the evidence.

There are two different kinds of evidence that support the ERH, one in which the studied material are weakly associated word-pairs and one in which the material consists of science passages. I will discuss each in turn. Carpenter and Yeung (2017) provided the first kind of evidence (see also, Carpenter, 2009; 2011; Rawson, Vaughn, & Carpenter, 2015). They had Ps study weakly associated word-pairs that varied in how strongly each pair's cue (i.e., the first word of the pair) was associated (pre-experiment) to its strongest associate. For example, one study on association norms established that the word *chalk* reminds people of its strongest associate, *board*, 69% of the time, whereas *soup* reminds people of its strongest associate, *chicken*, only 10% of the time (Nelson, McEvoy, & Schreiber, 2004). Carpenter and Yeung argued that the word-pair *chalk-crayon* would show a stronger testing effect than would the word-pair *soup-onion*. The reason for this is that Ps are more likely to think of an associate while being tested over the *chalk* word-pair (*chalk*-___?) than while being tested over the *soup* word-pair (*soup*-___?), for the strong associate that *chalk* has will likely require little time to become activated. Again, if an associate is activated, then its association to both members of the word-pairs will be strengthened, the number of retrieval routes to the second member of the pair will therefore be increased, and so the second member will now be more easily recalled when one is presented with the first member. Consistent with this reasoning, Carpenter and Yeung (2017) found that there was a bigger testing effect for words with stronger "strongest associates" than for words with weaker "strongest associates."

The other significant piece of evidence was given by Hinze et al. (2013). In their third experiment, they had Ps read a series of five texts, with each text covering a different scientific subject. Some Ps were then given time to reread each of the texts, whereas other Ps were asked to explain each of the texts without the benefit of having them available. All participants returned after seven days to take a test over the material. Ps who had been asked to explain the texts were better able to answer inference questions on those texts than were the participants who merely reread them. The authors concluded, in line with the ERH, that an attempt to remember what each text was about prompted participants to inter-relate various bits of information from within a text and from one's pre-existing background knowledge, and elaborate on how these bits of information fit together to form a coherent situation model. Attempts to remember the progression of a science text may have caused the memories of relevant background information to become activated, which is often essential for the integration process (Kintsch, 1994), and which often fails to occur when participants passively read a text (McNamara & Kintsch, 1996).

In contrast to the evidence supporting the ERH just given, Karpicke and Smith (2012) have provided evidence that seems to contradict the ERH, at least in its application to the learning of simple word-pairs. Specifically, they had Ps study Swahili-English word-pairs (e.g., wingu-cloud) and try to think of semantic mediators while doing so (e.g., *wingu* is orthographically similar to *wing*, and things with wings can fly up to a *cloud*). This task was meant to induce the kind of activity that the ERH postulates is responsible for the testing effect, but without having the activity occur during attempts to retrieve a specific memory. This task, though, resulted in no better memory for the word-pairs than did merely rereading the word-pairs. Hence, it seems that the generation and inter-association of mediators is not the process responsible for the testing effect. Carpenter and Yeung (2017) argued that the activation of mediators does not occur to the same extent in the task used by Karpicke and Smith (2012) as it does in retrieval attempts. Taken together, it is, at the very least, unclear if the ERH is adequate to explain the benefits of testing for the learning of simple material.

Furthermore, even though the ERH seems to be necessary to explain the benefit of testing for answering inference questions that Hinze et al. (2013) observed (see also, Butler, 2010; Johnson & Mayer, 2009; Karpicke & Blunt, 2011; McDaniel, Howard, & Einstein), it is still possible, as Butler (2010) pointed out, that testing only improves your ability to successfully remember information that you consciously try to remember. To answer an inference question, one must (1) recognize what information is relevant to answering the question, (2) successfully retrieve that information, and—because the actual answer to the question has not been explicitly stated prior—(3) successfully use or apply that information to solve the inference question

(Barnett & Ceci, 2002). I noted earlier that Kintsch's (1994) model suggests that testing will facilitate one's ability to complete step 1 (i.e., be reminded of the relevant information) if it facilitates the integration of that information into a well-constructed situation model. It seems clear that one will also be better able to complete step 3 if the relevant information is integrated within a situation model—that is, if one knows how the parts fit together. However, testing can still improve one's ability to solve inference problems, even if it does not facilitate construction of situation models (as seems to be suggested by the ECA), because testing will still improve one's ability to complete step 2, which is also necessary. Hinze et al.'s (2013) Ps may have therefore improved in their ability to answer inference questions despite testing not aiding the construction of their situation models (see further discussion of this issue in Butler, Black-Maier, Raley, & Marsh, 2017). In sum, considerations from the metacomprehension literature strongly suggest that tests over complex materials can aid construction of the situation model, at least if they are the right kinds of questions.

Testing Effects and Metacomprehension

Nguyen and McDaniel (2015) found that taking a test over a complex text improves one's ability to monitor how well one has truly understood that text, and to a degree greater than does note-taking. These results are important because this monitoring ability—i.e., metacomprehension—is generally quite meager (Maki, 1998), due both to overconfidence in general (i.e., poor absolute metacomprehension) and to a poor ability to specify which particular portions of a text are best (and least) understood (i.e., poor relative metacomprehension; Wiley et al., 2016). Faulty metacomprehension causes us to terminate reflection on the meaning of a text too early (because we are overconfident) and to inappropriately apportion our study efforts among the various portions of a text. The problem, though, is not that there is no effective way to monitor how well one has learned a text. Otherwise, Nguyen and McDaniel (2015) would not have found that tests cause metacomprehension accuracy to increase. Rather, students tend to use ineffective ways of monitoring their comprehension.

Griffin, Wiley, and Salas (2013) described the monitoring process within a cue-utilization framework. The premise of their metacomprehension model is that we are not able to directly gauge how well we understand a text. In contrast to cars, which have speedometers directly telling us how fast we are going, the mind does not provide a read-out that we, for example, understand 80% of a text. We must instead rely on less direct indicators Effective indicators are those that assess the coherence and completeness of the situation model that one has derived from a text, such as one's ability to self-explain a text either while reading it (Wiley et al., 2016) or with the text absent (Nguyen & McDaniel, 2015).

Thiede, Griffin, Wiley, and Anderson (2010) found that college students rarely use effective indicators, and are much more likely to base their metacomprehension judgments on how well they think that they have memorized the ideas explicitly stated in a text, on their familiarity with the topic, on their interest in the topic, or even on objective features of the text such as its length. Importantly, Thiede et al. (2010) also found that metacomprehension judgements based on these indicators were not nearly as accurate as were those based on indicators that tap the situation model. To sum up, tests foster elaboration of a given situation model—as postulated by the ERH—partially because they greatly facilitate one's ability to notice deficiencies in the situation model that need to be reflected upon and addressed (Nguyen & McDaniel, 2015). When information is thus well-integrated within a situation model, as opposed to being encoded in separate and unintegrated memory representations, we are better able to remember it and use it when trying to make sense of new related information processed at a later time. We should therefore predict a forward effect of testing in which testing improves our ability to understand related information that we process in the future. However, the kind of test questions being answered is very important.

Different Kinds of Test Questions

It is possible that taking a test can instill a memorization-processing approach, which would entail using monitoring cues that tap the textbase instead of the situation model. Such tests are therefore less likely to foster further construction of the situation model. Indeed, Hinze et al. (2013) found that asking Ps to *recall* as much information as they could from a prior text did not improve their ability to then answer inference questions about that text (at least not more so than did rereading). However, we have already stated that Hinze et al. (2013) did observe this improvement when Ps were asked to *explain* the prior text. The kinds of tests that are likely to improve learner's ability to assess their own understanding of a text are likely to be the same one's that allow researchers to assess how well they understand a text. One such kind of test is one which consists of inference questions (Kintsch, 2005). Unlike memory questions, which require Ps to remember something explicitly stated in a text, inference questions require Ps to infer what was only implicit in—or implied by—the text.

Most of the research on the testing effect have assessed the learning benefits of answering memory question rather than of answering inference questions. This is perhaps because most of the research on the testing effect have been interested in the benefits of *retrieving* information from memory. Karpicke and Zaromb (2010), for instance, differentiate the *testing effect* from the *generation effect* on the basis that only in the former case is one in "retrieval mode," or consciously trying to recall a prior experience. The *generation effect*, by contrast, refers to the fact that generating information produces better memory for that information than does reading it. For example, one will have a better memory for a list of words if he or she generates each word (e.g., "s _ _ rk") rather than reading each word (e.g., "shark"). Inference questions seem to be a combination of retrieval and generation, for one must both recall the relevant information and generate the inference.

There have been a few studies in which researchers tried to assess whether there are any benefits from answering inference questions rather than from answering memory questions. Butler et al. (2017) found that Ps were better able to abstract and apply concepts from a geology text—Ps had to use the content of multiple sentences to abstract a single one of these concepts—if they answered earlier questions that required them to apply these concepts than if they read additional text that gave examples of these concepts. Endres, Carpenter, Martin, and Renkl (2017) found that Ps developed a deeper understanding of a lecture if they were instructed to relate that information to episodes from their own lives during a subsequent free recall task than if they were merely asked to free recall that information. Jensen, McDaniel, Woodard, and Kummer (2014) found that students, over the semester of an actual course, could better apply the course content to novel, high level inference questions on the final exam if their prior quizzes consisted of inference questions than if they consisted of memory-retrieval questions (see also Glass, 2009; McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013). Lastly, and working in a slightly different tradition of research, Needham and Begg (1992) found that Ps were better able to solve a problem if they had earlier made inferences to solve an analogous problem than if they had been merely told the solution to that analogous problem (see also Darabi, Nelson, & Palanki, 2007). Given the above research, along with metacomprehension considerations, it seems likely that answering inference questions should result in a forward effect on comprehension compared to answering memory questions. However, to my knowledge, there are no studies that have assessed whether some formats of inference questions are more conducive to understanding than others.

I stated above that inferential questions will be particularly conducive to the elaboration of a situation model if they significantly improve the learner's metacomprehension of that situation model. From this, we can deduce that certain kinds of multiple-choice questions are likely better than are others for facilitating elaboration of the situation model. For example, easy multiple-choice questions are unlikely to be beneficial in this regard. Inferential multiple-choice questions can, however, easily alert learners to deficiencies in their situation model if each answer choice is accompanied by a sensible argument for it. Indeed, there is research showing the benefits of presenting correct information together with (and sometimes in dialogue with) incorrect information that contradicts it. In fact, the point of including this incorrect information is to confuse the learner, at least initially, for that confusion will drive further reflection (D'Mello, Lehman, Pekrun, & Graesser, 2014).

For example, D'Mello et al. (2014) had Ps listen to two computer personalities discuss concepts in research methods. Ps were better able to answer an inference question on a research methods concept if the two personalities argued while discussing the concept than if they were in agreement while discussing it. Similar studies have found this method helpful for teaching decimals (Durkin & Rittle-Johnson, 2012) as well as introductory physics (Muller, Bewes, Sharma, & Reimann, 2008). Inferential multiple-choice questions can function in a similar way if each answer choice is accompanied by an argument for it. There will then be a set of arguments contradicting each other, which will cause confusion if each argument seems, on the surface, reasonable. This confusion, of course, should be a strong metacomprehension indicator that the relevant situation model requires further elaboration. It seems that inference multiple-choice questions that do not provide arguments for each answer choice, along with inference short-answer questions, will be less effective because they do not present contradictory arguments.

Overview of the Experiments

The following experiments examine the forward effect of testing using inference questions. In particular, I am interested in determining if test questions over one set of information can improve one's ability to understand—and not just remember—new (and related) information that is presented after the test. It was earlier argued that such a forward effect will likely be obtained if there is a backward effect of testing. One's understanding of the subsequent chapters will be more facilitated if the test significantly improved one's understanding of the first chapter than if it only improved one's memory of it. Test questions can improve one's understanding of that text if they provide accurate metacomprehension feedback about parts of the respective situation model that are currently deficient or unspecified, for then one will be driven to reflect more about those parts of the situation model and remedy those deficiencies. Inference questions will likely give better metacomprehension feedback than will memory questions, and so result in a greater forward effect on comprehension. Similarly, inference multiple-choice questions should also have a greater forward effect of testing when each answer-choice is accompanied by an argument as opposed to when they are not.

The same basic procedure was used across four experiments. In the first two experiments, Ps began by reading a text describing the interference theory of forgetting. Half of the Ps then reread that text, whereas the other half took an inference test over that first text. All Ps then proceeded to read a text over the inhibition theory of forgetting, which is a more recent theory that in many ways built on interference theory. All Ps next took an inference test over both theories. Each question on this final test requires Ps to use and integrate information from both the first and the second text. It was hypothesized that the Ps who take a test over the first text would do better on the final test than would the Ps who reread the first text. It was also hypothesized that Ps in the Testing condition would report less mind-wandering than would Ps in the Rereading condition, as was found in Szpunar et al. (2014).

In the third experiment, there was no a rereading group. Rather, there was a group of Ps who took a memory test over the first text. It was hypothesized that the Ps who took an initial inference test would perform better on the final inference test than would Ps who took an initial memory test. In the final experiment, all Ps took an inference test over the first text. For half of the Ps, the inference test was multiple-choice and each answer was accompanied by an argument; the other Ps answered the same questions, but the answer-choices were not accompanied by arguments.   It was hypothesized that Ps would perform better on the final test if the initial test had arguments than if it did not.

Experiment 1

Participants

Ninety-one undergraduates (43 women) from the University of Oklahoma participated in Experiment 1. There were an additional 4 Ps who had to be excluded from analysis because they did not complete the study. Ps ranged from 18 to 47 years of age (M = 19.7). The Ps were predominantly White (73.6%), with the remaining Ps being American Indian or Native American (8.8%), Asian (6.6%), Black or African American (7.7%), Middle-Eastern (1.1%) and not reporting (2.2%). Ps received class credit for participating

Design

A 2-group (Rereading condition; Testing condition) between-subjects design was used.

Procedure

All Ps began by reading a text over the interference theory of forgetting (see Appendix A). The instructions at the beginning of this text informed them that they may be quizzed over the material in the text. Ps were then randomly assigned to either the Testing condition or the Rereading condition. Ps in the Rereading condition proceeded to reread the text over the interference theory of forgetting, whereas Ps in the Testing condition proceeded to take a 5-question test over that text. All Ps were next presented with the second text, which covered the inhibition theory of forgetting, and then took a 5-question test over it afterwards. Lastly, Ps answered questions about their mind-wandering throughout the study. There were no delays between any of these phases of the experiment.

Material

All material was presented to the Ps via the Qualtrics internet platform.

Texts. The text over the interference theory of forgetting was 2158 words in length and written at a twelfth-grade reading level according to Flesch-Kinkaid software. The text over the inhibition theory of forgetting was 1710 words in length and was also written at a twelfth-grade reading level according to Flesch-Kinkaid software.[1] Prior studies have used material of comparable length and difficulty (e.g., Butler, 2010; Jing et al., 2014; Kang, McDermott, & Roediger, 2007). Furthermore, a text needs to have a sufficient amount of complexity in order for there to be a significant distinction between its text-base and its situation model (Wiley et al., 2005). Each of the two texts was broken into segments, and Ps could only view one segment at a time. Ps had to view each segment for a certain amount of time before they could progress to the next segment. The time minima were, roughly, 15s less than the amount of time it takes to read the text aloud at a normal pace. The specific time minima are reported in Appendix A. When Ps in the Reread condition reread the first text, the time minima were retained and a 10-min time-limit was added.

Inference Questions. Two tests were constructed, one over the interference theory and one over the inhibition theory. The questions required Ps to infer components of these theories that were only implicit in the texts, and not explicitly stated. Such inference questions are necessary if one wants to assess how well Ps have constructed the situation model described in a text (Kintsch, 1994; Kintsch, 2005). Both tests consisted of five questions, which is sufficient to assess how well Ps have constructed the situation model described in a text (Weaver, 1990). The questions were multiple-choice, with four answer-choices per question. Each answer-choice was accompanied by a supporting argument. The supporting arguments were meant to cause confusion and metacognitive awareness of deficiencies of the Ps' understanding of the material being tested. Consequently, these accompanying arguments were meant to facilitate further construction of the situation model described in the texts. On each question, Ps had 12 "points" which they were to distribute among the answer-choices. If a P is certain that a particular choice is correct, then he or she should give all 12 points to that choice. However, if that P is uncertain which of two choices is correct, then he could give each choice 6 points. Using this "confidence-weighted testing" has been found increase the degree to which Ps consider all of the available choices (Sparck, Bjork, & Bjork, 2016). However many points a P allots to the correct answer of a given question was the number of points that he/she scored on that question[2]. Hence, the highest possible score for each test was 60. Ps were told the correct answer to each question immediately after having answered it. Lastly, Ps had to spend a certain minimum amount of time on each question. The time minima were roughly the amount of time it takes to read aloud the question and answers-choices at a normal pace. The specific time minima are reported in Appendix B.

Trick Question. In the final test, there was a trick question in addition to the five inference questions. Within the middle of the trick question, Ps were explicitly told "Pick the third answer." Failure to pick the third answer can serve as an indicator that the P was not fully reading the questions on the final test.

Mind-Wandering Questions. Ps were asked "How much did your mind wander while reading these texts on forgetting?" and "How much did your mind-wandering increase as you progressed through these texts on forgetting," with Likert scales ranging from 1 ("Not at all") to 7 ("Very Much"). See Szpunar et al. (2013) for a similar procedure. Additionally, I asked Ps how interesting the second text was relative to the first test, using a Likert scale ranging from -2 ("Much less interesting") to +2 ("Much more interesting").

Results

Average time spent on the experiment was 41 min. Ps in the Testing condition on average scored 26.7 out of 60 (SD = 11.9) on the test over the first text. There was no significant difference between the proportion of Ps in the Testing condition who correctly answered the trick question (.40, 95% CI [.26, .54]) and the proportion of Ps in the Rereading condition who correctly answered the trick question (.52, 95% CI [.37, .67]), z = 1.13, p = .26 (two-tailed). Similarly, there was no significant difference between the two groups with respect to time spent reading the second text, $t(89)$ = .38, p = .25. However, Ps in the Rereading condition did spend more time taking the final test (M = 747.0s, SD = 130) than did Ps in the Testing condition (M = 671.2s, SD = 130.1), $t(89)$ = 3.1, p = .003 (two-tailed, equal variances not assumed).

Contrary to my first hypothesis, Ps in the Testing condition did not score higher on the final test (*M* = 18.6; *SD* = 9.0) than did Ps in the Rereading condition (*M* = 19.1, *SD* = 11.2; see Figure 1). To further investigate the data, I scored Ps' test performances in two other ways. In both cases, I eliminated Ps' weightings. In the first case, Ps could get 1 point on a problem if they allotted more of their 12 points to the correct answer than to any of the incorrect answers; otherwise, they received a score of 0. Hence, across the five problems, the highest score Ps could achieve was a 5. I will refer to this measure as the number of problems answered correctly. Taking this approach, Ps in the Rereading condition again did not score lower (*M* = 1.0, *SD* = 1.2) than did Ps in the Testing condition (*M* = .90, *SD* = .87), which was contrary to my hypothesis (see Figure 2). My final way of scoring Ps' performances was to give them 1 point on each problem in which they did not give any evidence that they thought the correct answer was a good answer. Specifically, a P received 1 point on a problem if he/she ranked the correct choice as least likely (i.e., gave fewer points to that answer than to any of the other, incorrect answers), or if he/she evenly distributed the 12 points amongst the 4 answer-choices. A P therefore receives 1 point on a problem if he/she is "dead wrong," but otherwise receives 0 points. I will refer to this measure as the number of problems answered incorrectly. Ps in the Testing condition did not answer less problems incorrectly (*M* = 2.8, *SD* = 1.1) than did Ps in the Rereading condition (*M* = 2.3, *SD* = 1.3), which was again contrary to my hypothesis (see Figure 3).

The Ps' pattern of performance was not changed when analyses were limited to only the Ps who correctly answered the trick question. Ps in the Testing condition again did not score higher on the final test (*M* = 16.6, *SD* = 6.5) than did Ps in the Rereading condition (*M* = 19.8, *SD* = 10.8). Ps in the Testing condition also did not get more problems correct (*M* = 1, *SD* = .7) than did Ps in the Rereading condition (*M* = 1.2, *SD* = 1.2). Finally, Ps in the Testing condition did not get fewer problems incorrect (*M* = 3.1, *SD* = .8) than did Ps in the Rereading condition (*M* = 2.3, *SD* = 1.1).

Contrary to my second hypothesis, Ps in the testing group did not report less mind-wandering (*M* = 4.5, *SD* = 1.3) than did Ps in the rereading group (*M* = 4.8, *SD* = 1.5), $t(88) = .92$, $p = .36$, nor did they report less of an increase in mind-wandering as the study progressed (*M* = 5.2, *SD* = 1.3) than did Ps in the rereading group (*M* = 5.3, *SD* = 1.3), $t(88) = .92$, $p = .55$. On a related note, there was no significant difference between how interesting (relative to the first text) Ps in the testing condition found the second text (*M* = -.67, *SD* = 1.1) and how interesting Ps in the rereading condition found the second text (*M* = .82, *SD* = 1.1), $t(67) = .61$, $p = .54$. This data is presented in Table 1.

A final measure of interest is point-distribution, which is the average number of choices among which a P distributed his/her points on a given problem. For example, a P who allotted all 12 of his points to a single answer choice on each of the 5 problems would have a point-distribution score of 1. Across all 5 problems, Ps could distribute their points across a maximum number of 20 answer-choices (4 for each of the 5 questions). This measure can be indicative of various characteristics, such as how well the P knew the material or how much the P considered each of the alternatives. Determining how to interpret point-distribution scores is moot in this case, as there was no significant difference in point-distribution between the rereading condition (*M* = 2.6, *SD* = .95) and the testing condition (*M* = 2.5, *SD* = 1.1), $t(89) = .81$, $p = .42$.

Experiment 2

Experiment 2 was a replication of Experiment 1, with minor changes meant to address the fact that in Experiment 1, (a) Ps in the testing condition spent less time taking the final test than Ps in the rereading condition, suggesting greater fatigue in the testing condition than in the rereading condition, and (b) the proportion of responses to the trick question that were correct was very low, suggesting fatigue in general across the two groups. To address these issues, the texts were altered to be shorter and to have a lower reading level. The time minima on the texts were also reduced in order to reduce the amount of waiting time. The concern here was the Ps in Experiment 1 may have gotten bored waiting for the next button to appear. The trick question was also altered to be more explicitly a trick question. Additionally, each test question was changed to have only three answer-choices as opposed to four, thereby making them easier to answer. I again hypothesized that (1) the testing group would score higher on the final test than would the rereading group and (2) that the testing group would report less mind-wandering than would the rereading group.

Participants

Sixty-seven undergraduates (51 women) from the University of Oklahoma participated in Experiment 2. They ranged from 18 to 22 years of age (M = 18.5). The Ps were predominantly White (85.1%), with the remaining Ps being American Indian or Native American (6.0%), Asian (3.0%), Black or African American (4.5%), and not reporting (1.5%). Ps received class credit for participating.

There were 32 Ps randomly assigned to be in the Rereading condition and 35 Ps that were randomly assigned to be in the Testing condition. Ps were asked if they had previously studied the science of memory, using a Likert scale ranging from 0 (“Not at all”) to 4 (“Extensively”). There was no evidence that Ps in the Testing condition had studied the science of memory either more or less (*M* = 1.2, *SD* = .93) than had Ps in the Rereading condition (*M* = 1, *SD* = .87), *t*(65) = .93, *p* = .41.

Design

A 2-group (Rereading condition; Testing condition) between-subjects design was used.[3]

Procedures

The procedure was the same as Experiment 1.

Materials

All material was presented to the Ps via the Qualtrics internet platform.

Texts. The text over the interference theory of forgetting was reduced from the 2158 words in length down to 1546 words. The writing was also reduced from a twelfth-grade reading level down to a seventh-grade reading level, according to Flesch-Kinkaid software. The text over the inhibition theory of forgetting was actually increased from 1710 words to 1750 words. However, the writing was reduced from a twelfth-grade reading level down to a reading level of 8.6. The texts were again divided into segments. The time minimum for each segment was roughly 10s per paragraph. See Appendix C for the texts as well as the time minima.

Inference Questions. The inference questions were largely the same as those from Experiment 1, with the main difference being that each question now only had three answer-choices as opposed to four. Some questions were also made less difficult. Lastly, the time minima for the questions were lower than they were in Experiment 1. See Appendix D for the questions as well as the time minima.

Trick Questions. In the final test, there was a trick question in addition to the five inference questions. Within the middle of the trick question, Ps were explicitly told "This fourth question is a trick question, and I want you to give all twelve points to the third option."

Mind-Wandering Questions. The mind-wandering questions from Experiment 1 were retained. However, Ps were additionally asked "Have you previously studied the science of memory?" with the Likert scale ranging from 0 ("Not at all") to 4 ("Extensively").

Results

Ps on average spent 31 min on the experiment, which is roughly a .25 decrease from the average time spent in Experiment 1. Ps in the Testing condition on average scored 34.3 out of 60 (*SD* = 11.9) on the test over the first text. Additionally, .62 of the Ps correctly answered the trick question (95% CI [.51, .75]). The difference between the proportion of Ps in the rereading group who correctly answered the trick question (.72, 95% CI [.57, .87]) and the proportion of Ps in the testing group who correctly answered the trick question (.54, 95% CI [.37, .72]) approached significance, $z = 1.48$, $p = .068$. There was no evidence for a difference between the amount of time the Rereading condition spent on reading the second text (*M* = 487s, *SD* = 116) and the amount of time the Testing condition spent on the second text (*M* = 490s, *SD* = 150), $t(65) = -.09$, $p = .46$. However, Ps in the Rereading condition did spend more time on the final test (*M* = 607s, *SD* = 151.7) than did Ps in the Testing condition (*M* = 555s, *SD* = 122.4), $t(65) = 2.87$, $p = .002$ (two-tailed). Lastly, Ps in the Testing condition on average scored 34.2 out of 60 (*SD* = 11.9) on the test over the first text.

Contrary to my first hypothesis, Ps in the Testing condition (*M* = 21.7, *SD* = 9.1) did not score higher on the final test than did Ps in the Rereading condition (*M* = 24.2, *SD* = 11.2; see Figure 1). Ps in the Testing condition also did not get more problems correct (*M* = 1.31, *SD* = .90) than did those in the Rereading condition (*M* = 1.44, *SD* = 1.19; see Figure 2). Lastly, Ps in the Testing condition—again, contrary to my hypothesis—did not get fewer problems incorrect on the final test (*M* = 2.66, *SD* = 1.08) than did Ps in the Rereading condition (*M* = 2.06, *SD* = 1.01; see Figure 3).

The pattern of P performance was not changed when only Ps who correctly answered the trick question were taken into account. Ps in the Testing condition still did not score higher (*M* = 21.1, *SD* = 11.0) than did Ps in the Rereading condition (*M* = 24.7, *SD* = 10.9). Similarly, Ps in the Testing condition did not answer more questions correctly (*M* = 1.3, *SD* = 1.0) than did Ps in the Rereading condition (*M* = 1.5, *SD* = 1.1). Lastly, Ps in the Testing condition did not get fewer questions incorrect (*M* = 2.7, *SD* = 1.2) than did Ps in the Rereading condition (*M* = 2.0, *SD* = 1.0).

Contrary to my second hypothesis, Ps in the Rereading condition did not report more mind-wandering (*M* = 4.50, *SD* = 1.39) than did those in the Testing condition (*M* = 4.97, *SD* = 1.39), and they also did not report a greater increase in mind-wandering as the study progressed (*M* = 4.56, *SD* = 1.58) than did Ps in the Testing condition (*M* = 5.54, *SD* = 1.12). On a related note, there was no significant evidence that Ps in the testing group found the second text less interesting (*M* = -.22, *SD* = .98) than did Ps in the rereading group (*M* = -.37, *SD* = 1.1), $t(65) = .59$, $p = .56$. A succinct representation of this data is available in Table 2.

A final measure of interest was point-distribution, which was the average number of choices among which a P distributed his/her points on a given problem. Ps in the Testing condition distributed their points amongst fewer of the 15 answer-choices (3 for each of the 5 questions; *M* = 7.60, *SD* = 2.45) than did Ps in the rereading condition (*M* = 8.81, *SD* = 2.57), $t(65) = 1.98$, $p = .052$. Given the prior analyses already done, this most recent comparison suggests that Ps in the testing condition spent less effort comparing the plausibility of the various answer choices on a given problem.

Experiment 3

Experiment 3 was identical to Experiment 2 with the exception of two changes. First, the Rereading condition was dropped and replaced by a Memory Testing condition. Ps in this condition answered questions over the interference text after they first finished reading that text, as opposed to proceeding to read that text again. The test questions were in multiple-choice format and asked them about concepts explicitly described in the text. One potential issue is the fact that Ps in the Inference Testing condition were exposed to information that Ps in the Memory Testing condition were not, viz., the correct answers to the inference questions over the first text. Ps in the Inference Testing condition are exposed to this information regardless of whether they correctly answer the questions because they receive feedback immediately after answering each question. However, this benefit seems to be a natural extension of answering inference questions. Given that positive effects of answering questions are hard to detect (see Experiments 1-2), it did not seem desirable to eliminate this benefit of answering inference questions and thereby further increase the difficulty of detecting the possible positive effects of answering inference questions. The second change was that half of the Ps were told at the beginning of the experiment that their task was to memorize the texts as best they could, whereas the other Ps were told that their task was to understand the texts as best they could and that quizzes would test their understanding of the texts as opposed to their memory for the texts. There were two hypotheses in this experiment. The first hypothesis was that Ps in the Inference Testing condition would perform better on the final test than would Ps in the Memory Testing conditi on. The second hypothesis was that Ps in the Comprehension Instruction condition would score better on the final test than would Ps in the Memory Instruction condition. The logic driving this second hypothesis is that *test-expectancy* would drive Ps in the Comprehension Instruction condition to more closely read and think through, at the very least, the first text.

Participants

Ninety-six undergraduates (66 women) from the University of Oklahoma participated in Experiment 3. They ranged from 18 to 21 years of age (M = 18.6). The Ps were predominantly White (81%), with the remaining Ps being American Indian or Native American (3.1%), Asian (10.4%), Black or African American (3.1%), Native Hawaiian or other Pacific Islander (1.0%), and not reporting (1.0%). Ps received class credit for participating.

It is important to note that Ps who took the inference test reported a degree of prior exposure to the science of memory, on a scale from 0 to 4, that was not significantly different (M = 1.7, SD = .9) from that reported by the Ps who took the memory test (M = 1.5, SD = 1.0), $t(94)$ = 1.03, p = .31 (two-tailed).

Design

A 2(Inference Testing, Memory Testing) × 2(Comprehension Instruction, Memory Instruction) between-subjects design was used. There were 28 Ps in the Inference Testing condition who received the comprehension instruction, 23 Ps in the Inference Testing condition who received the memory instruction, 20 Ps in the Memory Testing condition who received the comprehension instruction, and 25 Ps in the Memory Testing condition who received the memory instruction.

Procedures

The procedure was similar to Experiment 1. The only difference was that, at beginning of the experiment, each P was told either:

> "Quizzes will test your **COMPREHENSION** of the presented information rather than your **MEMORY** of it. What this means is that the right answer to each question will have never been explicitly stated in the text; rather, you must understand the information well enough to be able to apply it to the quiz questions."

or “Quizzes will test your **MEMORY** for information explicitly stated in the texts.**”**

Material

All material was presented to the Ps via the Qualtrics internet platform.

Texts. The texts were the same as those used in Experiment 2.

Inference Questions. The inference questions were the same as those used in Experiment 2. The memory questions were in the same format as the inference questions. The only difference between the two types of questions was that the memory questions were answered directly in the texts whereas the inference questions were not (See Appendix D).

Trick Question. The trick question was the same as in Experiment 2.

Mind-Wandering Questions. The post-test questions from Experiment 2 were retained.

Results

Ps on average spent 29 min on the experiment (*SD* = 6.2). Ps in the Memory Testing condition on average scored a 36.2 out of 60 (*SD* = 13.0) on the first test; Ps in the Inference Testing condition on average scored a 24.3 out of 60 (*SD* = 11.3). Fifty-three percent of the Ps correctly answered the trick question (95% CI [.43, 63]). There was no significant evidence that Ps who took the memory test correctly answered the trick question at a higher frequency (.60, 95% CI [.45, .75]) than did Ps who took the inference test (.47, 95% CI [.33, .51]), $z = 1.27$, $p = .20$ (two-tailed). Ps who took a memory test spent more time reading the second text (*M* = 514.4s; *SD* = 105.3) than did Ps who took an inference test (*M* = 443.2s; *SD* = 126.8), $t(94) = 2.97$, $p = .004$ (two-tailed). However, there was no evidence that Ps who took the memory test spent a different amount of time on the final test (*M* = 485.9s; *SD* = 91.8) than did Ps who took the inference test (*M* = 496.0s; *SD* = 122.2), $t(94) = .45$, $p = .65$.

Contrary to my first hypothesis, Ps did not score higher on the final test when they took an initial inference test (*M* = 23.0, *SD* = 10.5) than when they took an initial memory test (*M* = 25.4, *SD* = 9.6; see Figure 1). Scoring Ps' final test scores in the alternative ways (outlined in the Results section of Experiment 1) did not reverse this trend (See Figures 2 and 3). Contrary to my second hypothesis, Ps did not score higher when given the comprehension instruction (*M* = 23.8, *SD* = 10.3), than when given the memory instruction (*M* = 24.5, *SD* = 10.0). There was also no significant evidence of an interaction between testing condition and instruction condition, F (1, 69) = 1.6, *p* = .21.

The Ps' pattern of performance looked largely the same when analyses were restricted to Ps who correctly answered the trick question. Ps in the Comprehension condition still did not score higher on the final test (*M* = 24.0, *SD* = 12.5) than did Ps in the Memory condition (*M* = 26, *SD* = 8.6). Similarly, Ps in the Comprehension condition still not get fewer questions wrong (*M* = 2.4, *SD* = 1.0) than did Ps in the Memory condition (*M* = 2.2, *SD* = 1.0). Ps in the Comprehension condition did *numerically* get more questions correct (*M* = 1.5, *SD* = 1.4) than did Ps in the Memory condition (*M* = 1.3, *SD* = 1.1), but this difference did not reach significance, *t*(49) = .7, *p* = .24.

There were no significant differences in interest of the second text [*t*(94) = 1.23, *p* = .22 (two-tailed)], in mind-wandering [*t*(94) = .04, *p* = .97 (two-tailed)], in increases in mind-wandering [*t*(94) = .96, *p* = .33 (two-tailed)], or in point-distribution [*t*(94) = .54, *p* = .59 (two-tailed)]. Table 3 gives the relevant statistics.

Experiment 4

Experiment 4 was identical to Experiment 3 except the Memory Testing condition was dropped and replaced by an alternative inference testing condition. In this new condition—the

Non-Argument Testing condition—Ps took an inference test over the first text, but the answer-choices on this test were not accompanied by supporting arguments. Thus, this test was different from the inference test over the first text that was used in Experiments 2 and 3. On that latter test, the answer-choices were accompanied by supporting arguments. Ps in the Argument Testing condition took this latter version of the test in Experiment 4. This difference with respect to the format of the inference test over the first text was the only point on which these two conditions differed. It was hypothesized that those in the Argument Testing condition would score better on the final test than would Ps in the Non-Argument Testing condition.

Participants

One hundred and eight undergraduates (85 women) from the University of Oklahoma participated in Experiment 4. They ranged from 18 to 23 years of age (M = 18.6). The Ps were predominantly White (75%), with the remaining Ps being American Indian or Native American (5.6%), Asian (9.3%), Black or African American (4.6%), Middle-Eastern (1.9%), and not reporting (3.7%). Ps received class credit for participating.

It is important to note that Ps in the Argument Testing condition test reported a degree of prior exposure to the science of memory, on a scale from 0 to 4, that was not significantly different (M = 1.2, SD = 1.0) from that reported by the Ps in the Non-Argument Testing condition (M = 1.3, SD = 1.1), $t(106) = .52$, $p = .61$ (two-tailed).

Design

A 2(Argument Testing, Non-Argument Testing) × 2(Comprehension Instruction, Memory Instruction) between-subjects design was used. There were 28 Ps in the Non-Argument Testing condition who received the comprehension instruction, 27 Ps in the Non-Argument Testing condition who received the memory instruction, 27 Ps in the Argument Testing condition who received the comprehension instruction, and 26 Ps in the Argument Testing condition who received the memory instruction.

Procedure

The procedure was the same as Experiment 3.

Material

All material was presented to the Ps via the Qualtrics internet platform.

Texts. The texts were the same as those used in Experiment 2.

Test Questions. The questions over the first text for Ps in the Argument Testing condition were the same as those used in Experiment 2. The questions over the first text for Ps in the Non-Argument Testing condition were similar to those presented in the Argument Testing condition. However, the answer-choices were not accompanied by arguments. Also, in some cases the answer-choices were slightly altered, although questions themselves were always the same.

The reason why some of the answer-choices had to be altered had to do with the fact that, on the test with accompanying arguments, the only difference between some answer-choices was a difference in supporting argument. That is, two answer-choices sometimes had the *same* response, but the *argument* for one was inaccurate, and so the P had to choose the answer-choice that had both the correct response and the same argument. Because there were no arguments present for Ps in the Non-Argument condition, there could be no two answer-choices for a given question that had the same response. Hence, some of the answer-choices had to be changed for the Non-Argument test. Importantly, Ps were shown the argument proving the correct answer as feedback immediately after answering each question. (See Appendix D).

Trick Question. The trick question was the same as in Experiment 2.

Mind-Wandering Questions. The post-test questions from Experiment 2 were retained.

Results

Ps on average spent 24 min on the experiment (*SD* = 5.4). Sixty-five percent of Ps correctly answered the trick question (95% CI [.56, .74]). There was not significant evidence that Ps who took the test with arguments correctly answered the trick question at a different frequency (.65, 95% CI [.52, .78]) than did Ps who took the inference test without arguments (.66, 95% CI [.53, .79]), $z = .09$, $p = .93$ (two-tailed).

I ran a 2(Test: No-Argument, Argument) × 2(Instruction: Memory, Comprehension) between-subjects analysis of variance (ANOVA), with test performance as the dependent variable. There was some support for the primary hypothesis of Experiment 4. Specifically, Ps did score marginally higher when the answer-choices of the initial test had arguments (*M* = 25.0, *SD* = 10.3) than when the answer-choices did not have arguments (*M* = 22.0, *SD* = 11.0), F (1, 104) = 2.1, $p = .077$, Cohen's $d = .28$ (see Figure 1). I did not detect a difference in performance between Ps given the memory instruction (*M* = 24.8, *SD* = 10.9) and Ps given the comprehension instruction (*M* = 24.5, *SD* = 10.9), F (1, 104) = .05, $p = .82$. Finally, I did not detect an interaction between the Test condition and the Instruction condition, F (1, 104) = .26, $p = .61$.

Alternative scoring did result in further support for the primary hypothesis. Ps in the Argument Testing condition got marginally more questions correct (*M* = 1.4, *SD* = 1.2) than did Ps in the Non-Argument Testing condition (*M* = 1.1, *SD* = 1.2), F (1, 104) = 2.3, $p = .068$, Cohen's $d = .29$ (see Figure 2). Furthermore, Ps in the Argument Testing condition got significantly fewer problems incorrect (*M* = 2.3, *SD* = 1.0) than did Ps in the Non-Argument Testing condition (*M* = 2.7, *SD* = 1.3), F (1, 104) = 3.1, $p = .04$, Cohen's $d = .35$, (see Figure 3).

Ps who took an initial test with arguments did not differ from the other Ps with respect to time spent reading the second text [$t(102) = .76$, $p = .45$, (equal variance not assumed)], to time spent taking the final test [$t(106) = 1.4$, $p = .15$], to mind-wandering [$t(106) = .71$, $p = .48$], to prior knowledge [$t(106) = .52$, $p = .61$], to interest of the second text in comparison to the first text [$t(106) = .88$, $p = .38$], or to an increase in mind-wandering [$t(106) = 1.8$, $p = .069$] (see Table 4). Additionally, there also seemed to be marginal evidence that Ps in the Argument Testing condition distributed their points less on the final test ($M = 8.1$, $SD = 2.3$) than did Ps in the Non-Argument Testing condition ($M = 9.2$, $SD = 3.0$), $t(106) = 2.12$, $p = .036$. However, due to a Bonferroni correction that would lower the α-threshold to around .007, it seems prudent to take this final analysis with caution.

Discussion

In the current study, I tested a number of predictions about forward effects of testing. First, I hypothesized that taking an inference test over a text would result in better learning of a subsequent related text than would rereading that first text. There were two primary reasons why I hypothesized that this would happen. First, it has been found that taking a test improves one's understanding of the material to a greater degree than does rereading that material (Hinze, et al., 2013). Second, W. Kintsch's (1994) theory of comprehension states that one better comprehends a text when one has more background knowledge. Thus, taking a test over a text should improve one's ability to use that information as background knowledge when trying to learn a related topic. I also hypothesized that taking a test over a text would result in less subsequent mind-wander than would rereading that text, in accordance with prior research (Szpunar et al., 2013). My third hypothesis was that taking an inference test would result in a larger forward effect than would taking a memory test. This hypothesis is suggested by the fact that inference tests have been found to result in a better understanding of the tested material than have memory tests (Hinze et al., 2013) and because answering inference questions should result in better feedback regarding deficiencies in one's understanding of the text than should memory questions (Thiede et al., 2010). Lastly, I predicted that a greater forward effect of testing would be obtained when each of the answer-choices of the test are accompanied by an argument than when the answer-choices are not accompanied by an argument. Again, I assumed that answer-choices accompanied by arguments would be more metacognitively informative (and confusing, D'Mello et al., 2014) than would be answer-choices unaccompanied by arguments.

The results of the current study were, for the most part, not as predicted. Answering inference questions over a text did not result in better learning of a subsequent related text than did either rereading the first text (Experiments 1-2) or taking a memory test over the first text (Experiment 3). However, answering inference questions in which each answer-choice was accompanied by an argument did result in marginally better understanding of the second text than did answering inference questions in which the answer-choices were not accompanied by an argument (Experiment 4). Thus, Ps in 3 of the experiments did not seem to benefit from taking inference tests.

One explanation for the null results is that testing may mainly benefit the learning process by updating retrieved information with context elements, and *not* by fostering the integration of different bits of information. This possibility is predicted by the ECA (Karpicke et al., 2014). However, the ECA was developed almost exclusively in the context of how testing improves *memory*, with *comprehension* receiving scant attention. Furthermore, other research seems to strongly suggest (1) that testing has metacognitive benefits and (2) that such metacognitive benefits result in better learning (e.g., Nguyen & McDaniel, 2015). Thus, this explanation of the result does not seem likely.

A more likely explanation of the results has to do with the motivation of the Ps. The Ps in these experiments were in no significant way required to learn the material well. They merely had to finish their trial in order to receive course credit. Hence, one of the major goals that drives student learning—receiving a high grade—was not present. Goals are a very important factor in the learning process though, and setting specific goals has been found to increase effort (Bryan & Locke, 1967). It seems that the effort of the Ps could have been better considering that the proportion of them who correctly answered the trick question ranged from 40-72% across the four experiments. It should be noted, though, that interpretation of responses to the trick question is not as simple as one would like. For example, in Experiment 4, the Ps in the Argument Testing condition who correctly answered the trick question performed worse on the final test than did Ps in the Argument Testing condition who did not correctly answer the trick question. This was in part because the two Ps who scored highest on the final test did not correctly answer the trick question. By contrast, the only P across all 4 experiments to score 0 points on the final test did happen to answer the trick question correctly. It cannot be gainsaid, though, that the Ps in the current study could have put forth greater effort and that this could have been facilitated by helping them to form goals (e.g., perform well enough to win a prize).

Setting a specific goal is also associated with greater persistence (LaPorte & Nath, 1976). Persistence seemed to be an issue for the inference-testing group in particular, with those taking an inference test over the first text spending less time on the final test than those who reread the first text (Experiments 1 and 2) and less time reading the second text than did Ps who took a memory test over the first text (Experiment 3). One interpretation of these low reading times is that the Ps were simply trying to complete the experiment and were not worried about spending extra time trying to understand the content. Perhaps, because taking an inference test is difficult, Ps must have persistence in order to benefit from taking that test. However, the Ps in Experiment 4 complicate interpretation of reading times and test times. The Argument Testing condition in Experiment 4 was the exact same as the Testing condition in Experiment 2 and the Inference Testing condition in Experiment 3. Yet, the Ps in the Argument Testing condition of Experiment 4 not only scored numerically better on the final test than did their counterparts from Experiments 1 and 2; they also spent even *less* time on reading the second text and on taking the final test (see Tables 2-4). It is still likely, nonetheless, that Ps are more likely to have their learning facilitated by test-taking when they have an actual *goal* of learning.

On a related note, interest has also been found to be an important factor in the learning process. For example, students with a greater interest in a subject are better able to answer inference questions in that subject than are students with less interest (Schiefele, 1992, Experiment 1). Students with more interest in a topic also tend have higher levels of energy, to make more elaborations (e.g., comparing information in the topic to personal experience or generating content-related images), and to experience greater amounts of flow (Schiefele, 1992, Experiment 2).

From Experiment 1 to Experiment 2, I tried to improve the interest of the texts in a couple ways. First, I reduced the time minima accompanying each text section and each test question. The relatively lengthy time minima used in Experiment 1 were meant to stimulate additional thinking. However, after I observed a null-result in Experiment 1, I assumed that Ps likely were spending that additional time merely waiting for the “Next” button to appear, with that waiting resulting in boredom and disinterest. Second, I reduced the length of the first text.

Third, I significantly reduced the reading level of the texts. It was hoped that this would lower the likelihood of a P becoming disengaged from the task due to its difficulty. Finally, I also tried to make the content of the texts somewhat more engaging. Another possible avenue explorable by future studies is figuring out which study domains are generally considered to be highly interesting. (An analysis of text-dependency was also started. See Appendix E for preliminary results).

Of course, a given text will not seem equally interesting to everybody. The Ps in Experiment 4 may in fact have found the materials used in the current study to be more interesting than did Ps in the first 3 Experiments. Again, the Argument Testing condition in Experiment 4 was the *exact* same as the Testing condition in Experiment 2 and the Inference Testing condition in Experiment 3. Nonetheless, Ps in the Argument Testing condition in Experiment 4 scored numerically better than those respective Ps in Experiments 2 and 3. (Of course, cross-experiment statistical analyses cannot be conducted, and so we cannot know if these differences were likely due merely to chance). At the same time, Ps in the Argument Testing condition of Experiment 4 reported numerically *less* prior exposure to the science of memory than did the respective Ps from Experiments 2 and 3. It seems unlikely therefore that the Ps in Experiment 4 had more experience with the science of memory. On the other hand, the Ps in the Argument condition of Experiment 4 also reported numerically *less* of an increase in mind-wandering than did the respective Ps from Experiments 2 and 3, which suggests that the Ps in Experiment 4 may have found the material to be more interesting (see Tables 2-4). Ps in Experiment 4 may have benefited from the argument-included inference tests because they had enough interest to (a) work on the deficiencies in their understanding of the first text that became apparent during their test over that text and (b) try to understand the more difficult second text.

It has already been suggested above that one way to improve the effort of Ps in studies similar to the current one is to choose topics that are more popular. Another suggestion might be to break up the study into multiple days. Answering inference questions is difficult (Jensen et al., 2014), and it is therefore not surprising (in retrospect) that the group that had to answer inference questions either spent less time reading the second text (Experiment 3) or taking the final test (Experiments 1 and 2) than did the comparison group. Answering the inference questions seems to have fatigued Ps, which is why splitting the experiment up across multiple days may be a good idea, for that would allow the Ps to rejuvenate.

The current study was conducted mainly out of pedagogical concerns. Many researchers have argued recently for an increase in the use of tests in education (e.g., Dunlosky et al., 2013), but it is important to consider what kinds of tests are most beneficial to students. I have argued that tests which consist primarily of memory questions are problematic because they result in untoward *test-expectancy effects* (McDaniel et al., 1994), such that students spend their study time trying to maximize the number of details memorized rather than trying maximize their general understanding of the topic. In fact, it seems that students may have confusion about whether or not there is any difference between memory and understanding (Jensen et al., 2014). Thus, it seems imperative that inference questions be used on tests with a relatively high degree of frequency. The question, though, is whether certain kinds of inference questions are more beneficial than others. Results from Experiment 4 suggests that inference-based multiple-choice question have a greater forward-effect on future learning when each answer-choice is accompanied by an argument. Nonetheless, much further research into this topic must be done.

References

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? *Psychological Bulletin*, *128*, 612–637.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417-444. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1146/annurev-psych-113011-143823

Bryan, J. F., & Locke, E. A. (1967). Goal setting as a means of increasing motivation. *Journal of Applied Psychology*, *51*(3), 274-277. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1037/h0024566

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1118–1133.

Butler, A. C., Black-Maier, A., Raley, N. D., & Marsh, E. J. (2017). Retrieving and applying knowledge to different examples promotes transfer of learning. *Journal of Experimental Psychology: Applied*, *23*(4), 433-446. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1037/xap0000142

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1563–1569.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1547–1552.

Carpenter, S. K., & Yeung, K. L. (2017). The role of mediator strength in learning from retrieval. *Journal of Memory and Language*, *92*, 128-141.

doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1016/j.jml.2016.06.008

Darabi, A. A., Nelson, D. W., & Palanki, S. (2007). Acquisition of troubleshooting skills in a computer simulation: Worked example vs. conventional problem solving instructional strategies. *Computers in Human Behavior, 23*(4), 1809-1819. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1016/j.chb.2005.11.001

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*, 153-170. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1016/j.learninstruc.2012.05.003

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1177/1529100612453266

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22*(3), 206-214. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1016/j.learninstruc.2011.11.001

Endres, T., Carpenter, S., Martin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting. *Learning and Instruction, 49*, 13-20. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1016/j.learninstruc.2016.11.010

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review, 62*(5), 369-377. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1037/h0046888

Glass, A. L. (2009). The effect of distributed questioning with varied examples on exam performance on inference questions. *Educational Psychology, 29*, 831– 848.

http://dx.doi.org/10.1080/0144341090 3310674

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*(3), 306-355. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1016/0010-0285(80)90013-4

Griffin, T. D., Wiley, J., & Salas, C. (2013). Supporting effective selfregulated learning: The critical role of monitoring. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 19 –34). New York, NY: Springer Science. http://dx.doi.org/10.1007/978-1-4419-5546-3_2

Hinze, S. R., Wiley, J., & Pellegrino, J. W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory and Language*, *69*, 151–164. http://dx.doi.org/10.1016/j.jml .2013.03.002

Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test. . . or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, *26*, 307–329. http://dx.doi.org/10.1007/s10648013- 9248-9

Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture. *Journal of Experimental Psychology: Applied*, *22*, 305–318. http://dx.doi.org/10.1037/xap0000087

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, *101*, 621– 629. http://dx .doi.org/10.1037/a0015183

Kang, S. H. K., McDermott, K. B., & Roediger, Henry L., I.,II. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*(4-5), 528-558.

doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1080/09541440601056620

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

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. *Psychology of Learning and Motivation*, *61*, 237–284. http://dx.doi.org/10.1016/B978-0-12-800283- 4.00007-1

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, *67*, 17–29.

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language, 62*(3), 227-239. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1016/j.jml.2009.11.010

Kintsch, E. (2005). Comprehension theory as a guide for the design of thoughtful questions. *Topics in Language Disorders, 25*(1), 51-64. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1097/00011363-200501000-00006

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*, 294 –303. http://dx.doi.org/10.1037/0003-066X.49.4 .294

LaPorte, R. E., & Nath, R. (1976). Role of performance goals in prose learning. *Journal of Educational Psychology, 68*(3), 260-264. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1037/0022-0663.68.3.260

Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117–144). Hillsdale, NJ: LEA.

Mannes, S. M., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and Instruction*, *4*(2), 91-115. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1207/s1532690xci0402_2

McDaniel, M. A., Blischak, D. M., & Challis, B. (1994). The effects of test expectancy on processing and memory of prose. *Contemporary Educational Psychology*, *19*(2), 230-248. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1006/ceps.1994.1019

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read- recite-review study strategy: Effective and portable. *Psychological Science*, *20*, 516–522.

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*(3), 360-372. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1002/acp.2914

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*(3), 247-288. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1080/01638539609544975

Mensink, G. M., & Raaijmakers, J. G. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology*, *33*(2), 172-186. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1016/0022-2496(89)90029-1

Momsen, J. L., Offerdahl, E., Kryjevskaia, M., Montplaisir, L., Anderson, E., & Grosz, N. (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. *CBE - Life Sciences Education*, *12*, 239–249.

Muller, D. A., Bewes, J., Sharma, M. D., & Reimann, P. (2008). Saying the wrong thing: Improving learning with multimedia by including misconceptions. *Journal of Computer*

Assisted Learning, 24(2), 144-155. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1111/j.1365-2729.2007.00248.x

Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. *Memory & Cognition, 19*(6), 543-557. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.3758/BF03197150

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*, 402–407.

Nguyen, K., & McDaniel, M. A. (2016). The JOIs of text comprehension: Supplementing retrieval practice to enhance inference performance. *Journal of Experimental Psychology: Applied*, *22*(1), 59-71. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1037/xap0000066

Nunes, L. D., & Weinstein, Y. (2012). Testing improves true recall and protects against the build-up of proactive interference without increasing false recall. *Memory, 20*(2), 138-154. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1080/09658211.2011.648198

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning* (NCER 2007–2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Pastötter, B., & Bäuml, K. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology, 5*, 1-5.

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

Rawson, K. A., Vaughn, K. E., & Carpenter, S. K. (2015). Does the benefit of testing depend on lag, and if so, why? Evaluating the elaborative retrieval hypothesis. *Memory & Cognition*, *43*, 619–633.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249 –255. http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., III, Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. *Psychology of Learning and Motivation*, *44*, 1–36.

Roediger, Henry L., I.,II, & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition, 1*(4), 242-248. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1016/j.jarmac.2012.09.002

Sahakyan, L., & Hendricks, H. E. (2012). Context change and retrieval difficulty in the list-before-last paradigm. *Memory & Cognition, 40*(6), 844-860. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.3758/s13421-012-0198-0

Schiefele, U. (1992). Topic interest and levels of text comprehension. In K. A. Renninger, S. Hidi & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 151-182) Lawrence Erlbaum Associates, Inc, Hillsdale, NJ.

Sparck, E., Bjork, E., & Bjork, R. (2016). On the learning benefits of confidence-weighted testing. *Cognitive Research: Principles and Implications, 1*(1), 1-10.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy* *of Sciences of the United States of America*, *110*, 6313– 6317. http://dx.doi.org/10.1073/pnas .1221764110

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1392–1399. http://dx.doi.org/10.1037/a0013082

Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1267–1280.

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes*, *47*, 331–362.

Thiede, K. W., Wiley, J., & Griffin, T. D. (2011). Test expectancy affects metacomprehension accuracy. *British Journal of Educational Psychology*, *81*, 264 –273. http://dx.doi.org/10.1348/135910710X510494

Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning & Verbal Behavior, 13*(2), 181-193. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.1016/S0022-5371(74)80043-5

Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information. *Memory & Cognition, 43*(1), 27-38. doi:http://dx.doi.org.ezproxy.lib.ou.edu/10.3758/s13421-014-0455-5

Weaver, C. A. (1990). Constraining factors in calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 214 –222.

Wiley, J., Griffin, T. D., Jaeger, A. J., Jarosz, A. F., Cushen, P. J., & Thiede, K. W. (2016). Improving metacomprehension accuracy in an undergraduate course context. *Journal of*

Experimental Psychology: Applied, 22(4), 393-405.

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review, 18*, 1140 –1147. http://dx.doi.org/10 .3758/s13423-011-0140-7

Yang, C., Potts, R., & Shanks, D. R. (2017). The forward testing effect on self-regulated study time allocation and metamemory monitoring. *Journal of Experimental Psychology: Applied*, *23*(3), 263-277.

Endnotes

[1]Flesch-Kindaid software calculates the reading level of a text with an equation that takes into account average sentence length and average word length. See https://www.webpagefx.com/tools/read-able/flesch-kincaid.html.

[2]One issue that was noticed only on the penultimate day of testing was that a P can, in some circumstances, see how many points prior Ps assigned to each answer-choice. It is not clear how this would affect performance, as each answer-choice was assigned many points by some prior Ps and few points by others.

[3]In Experiment 2, Ps were to receive either the Memory instruction or the Comprehension instruction, as described in the methods section of Experiment 3. However, due to a programming error, all of the first half of the Ps only received one of these two instructions. Thus, the instruction variable was not taken into account in the analyses of Experiment 2. It is important to note that the instruction variable was in no way confounded with the testing variable.

Tables

Table 1

Descriptive Statistics for Experiment 1

Condition	*n*	Time on Text 2	Time on Final Test	Mind-Wandering	Increase in Mind-Wandering	Interest in Text 2	Point Distribution
Rereading	44	663.0s (62.1)	747.0s (130.1)	4.8 (1.5)	5.3 (1.3)	-.82 (1.06)	10.5 (3.8)
Testing	47	677.2s (88.5)	671.2s (100.0)	4.6 (1.3)	5.2 (1.3)	-.67 (1.08)	9.9 (4.2)

Note. The standard deviations are provided within the parentheses.

Table 2

Descriptive Statistics for Experiment 2

Condition	*n*	Time on Text 2	Time on Final Test	Mind-Wandering	Increase in Mind-Wandering	Interest in Text 2	Prior Knowledge	Point Distribution
Rereading	32	486.7s (114.5)	607.2s (151.7)	4.5 (1.4)	4.6 (1.6)	-.22 (1.0)	1.0 (1.0)	8.9 (2.6)
Testing	35	489.8s (147.9)	555.2s (122.4)	5.0 (1.3)	5.5 (1.1)	-.37 (1.1)	1.2 (.9)	7.6 (2.5)

Table 3

Descriptive Statistics for Experiment 3

Condition	*n*	Time on Text 2	Time on Final Test	Mind-Wandering	Increase in Mind-Wandering	Interest in Text 2	Prior Knowledge	Point Distribution
Memory	45	514.4s (105.3)	485.9s (90.8)	4.9 (1.4)	5.3 (1.4)	-.31 (1.2)	1.5 (1.0)	8.4 (2.7)
Comprehension	51	443.2s (125.6)	496.0s (121.0)	4.9 (1.5)	5.5 (1.5)	-.61 (1.1)	1.7 (.9)	8.7 (2.8)

Table 4

Descriptive Statistics for Experiment 4

Condition	*n*	Time on Text 2	Time on Final Test	Mind-Wandering	Increase in Mind-Wandering	Interest in Text 2	Prior Knowledge	Point Distribution
Non-Argument	55	437.3s (126.1)	508.3s (142.4)	4.7 (1.6)	5.4 (1.4)	-.47 (1.1)	1.25 (1.1)	9.2 (3.0)
Argument	53	421.7s (100.5)	475.5s (87.8)	4.5 (1.6)	4.8 (1.7)	-.28 (1.2)	1.15 (1.0)	8.1 (2.3)

Figures

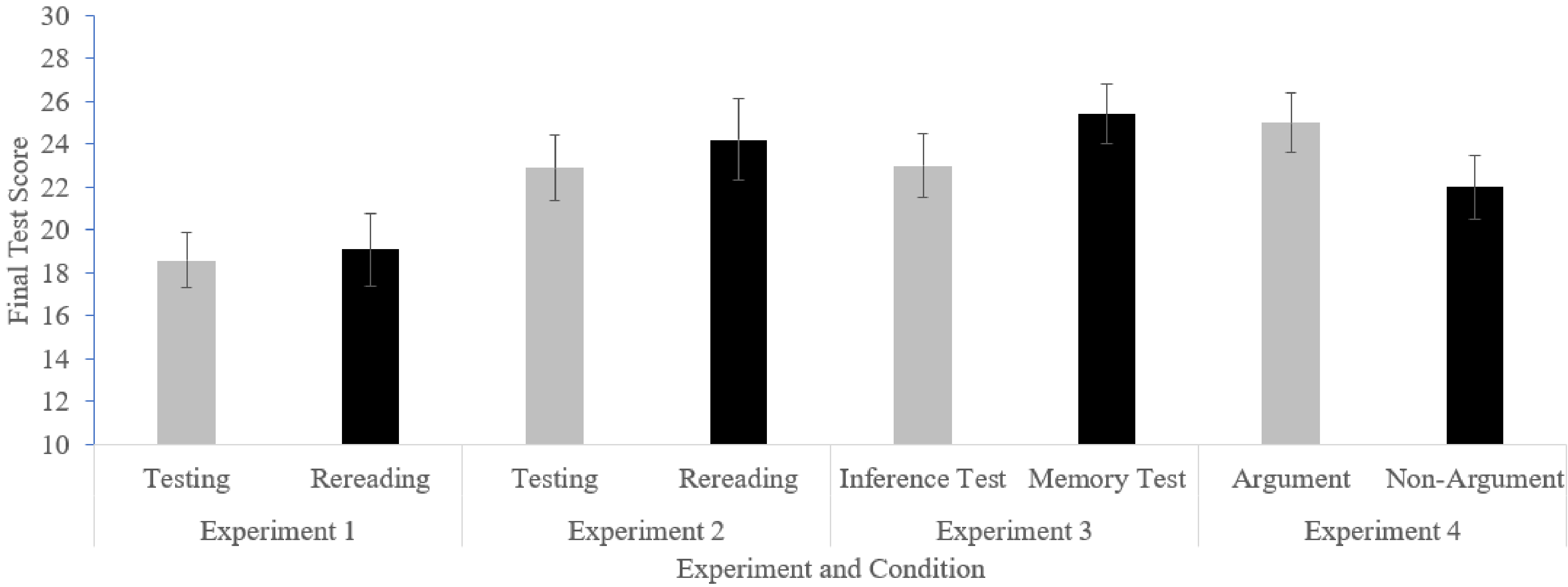

Figure 1. Average scores on the final test are shown. With this scoring method, Ps earned points according to how many of their 12 points they allotted to the correct answer. If a P allotted 7 of his/her 12 points to the correct answer on a given question, then 7 points were earned for that question. Across the 5 questions on the final test, Ps could earn a maximum total of 60 points.

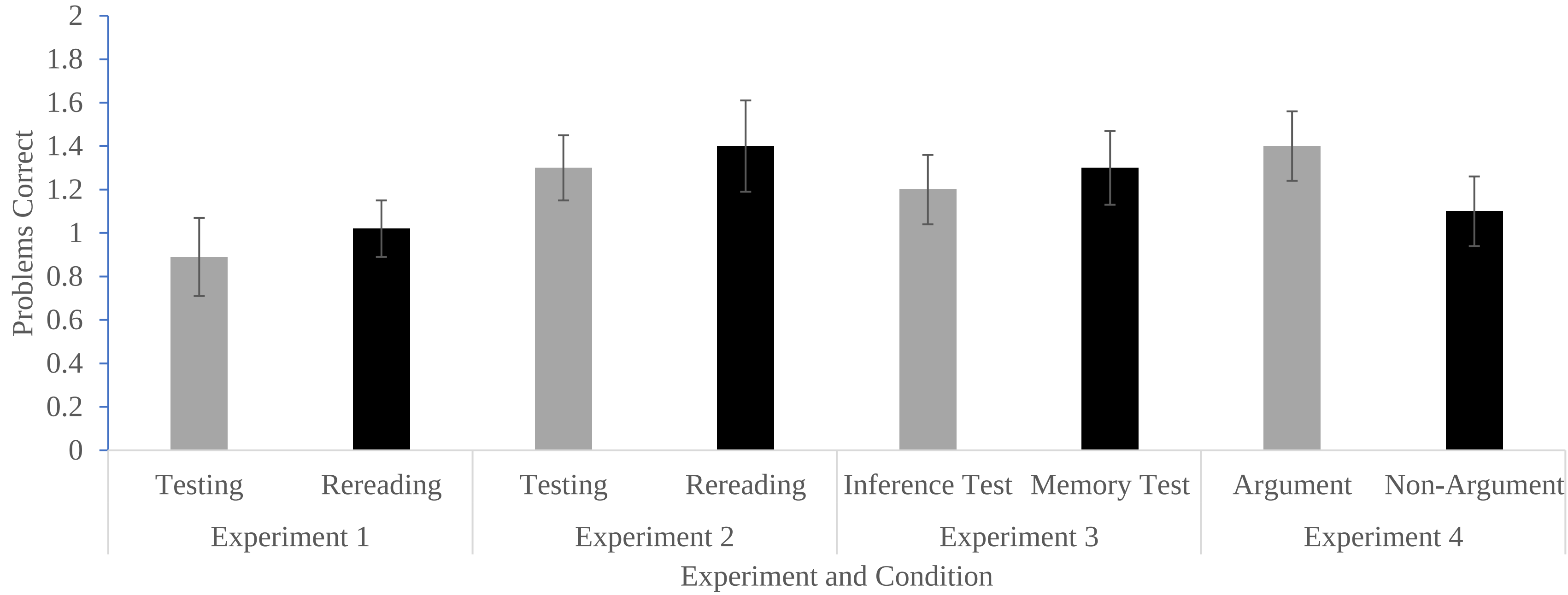

Figure 2. Presented is the average number of problems answered correctly on the final test. With this scoring method, a P was said to have correctly answered a question if the correct answer—relative to the other answer-choices—received the greatest proportion of the P's 12 points. The maximum possible score was a 5.

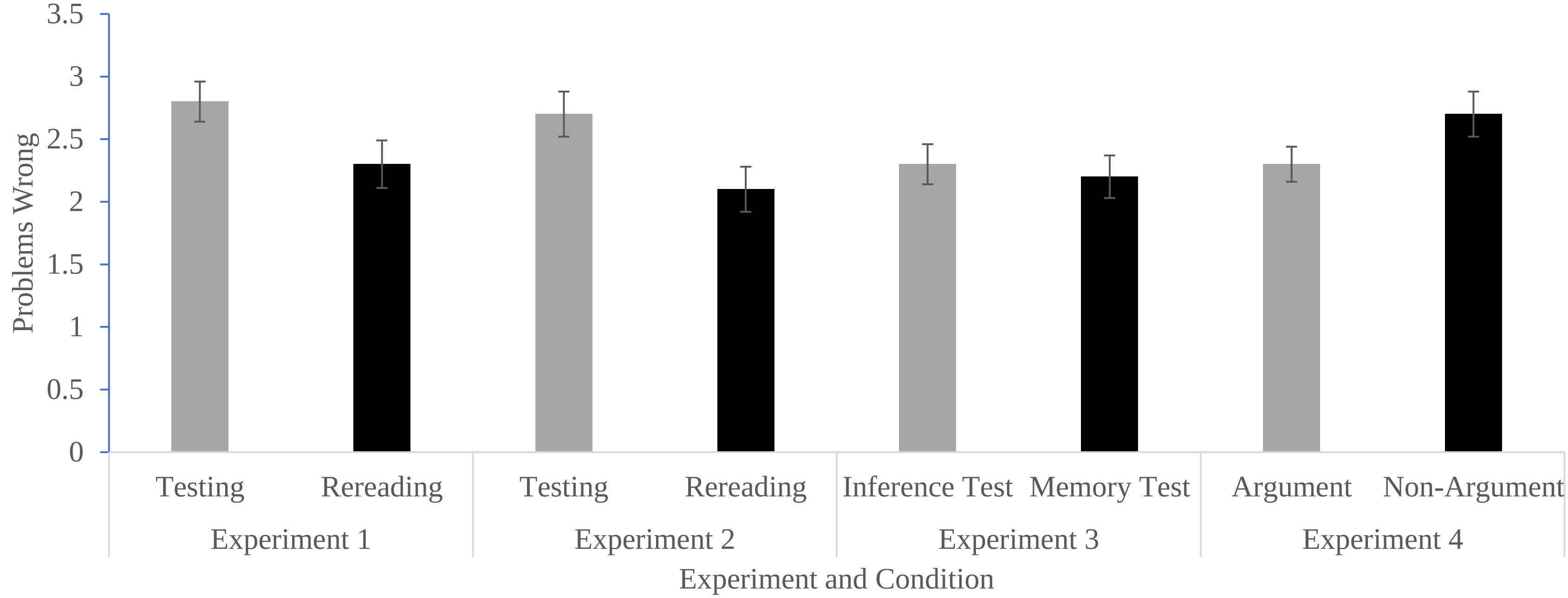

Figure 3. Presented is the average number of problems answered incorrectly on the final test. With this scoring method, a P was said to have incorrectly answered a question if the correct answer—relative to the other answer-choices—received the smallest proportion of the P's 12 points or if the P even distributed the 12 points amongst all the answer-choices. The worst possible score was a 5.

Appendix A

Experiment 1: Text 1

Section 1 (Minimum time = 7s)

Everyone is familiar with the process of remembering, with students in particular being frustrated by it and its limitations. Nevertheless, scientists are beginning to discover that such forgetting is possibly adaptive; they are of course also learning that the processes of remembering and forgetting have features that no one would have predicted. This article will cover these features and crucial experiments that have shed light on these questions.

Section 2 (Minimum time = 25s)

There are a number of basic assumptions made in the science of memory. For example, it is assumed that the act of remembering is a cue-driven process. This means that when you remember, say, that you have been meaning to call your mother, you do not just remember this out of the blue. Rather, you were thinking about something else which reminded you to call your mother. For example, a colleague may have been talking to you about his or her own mother, and that served as a reminder (or as a "cue"). Or perhaps that colleague brings a Red Bull into work that day, and this reminds you to call your mother, for your mother drinks Red Bull every day. In an event, for you to remember to call your mother, you must first be thinking about something else that reminds you to call her.

Section 3 (Minimum time = 40s)

It should be noted that the cue which reminds you of some memory must have a connection to that memory in your mental system. Thinking about the foundation plan of a house, for example, is unlikely to make you think about milk, nor is thinking about the cookies you're eating likely to remind you of different possible roof plans. However, thinking about the cookies you're eating can clearly remind you of milk. This is because you have some associative link connecting your mental representations (or memories) of milk and of cookies. However, most of us have not created an association between cookies and roof plans, and so thinking about the one will not cause us to think about the other (that is, unless you developed a habit of eating cookies on the roof so that, like an adolescent smoker, you can avoid being caught by your family). Now consider how thinking about Oklahoma might remind us of its NBA team (the Thunder), of its best restaurant (Victoria's Pasta), or of tornadoes. We can only remember one thing at a time. So, what will we be reminded of first whenever we think about Oklahoma?

Section 4 (Minimum time = 60s)

Psychologists have classically assumed that 2 factors determine the course of the memory-retrieval process: <u>association strength</u> and <u>trace strength</u>. ("Trace" is here being used in the sense of a "memory trace," as when we say that we have information about water stored in one memory trace and information about fire stored in a different memory trace). The concepts of association strength and trace strength can be illustrated by thinking about Arnold Schwarzenegger and Ronald Reagan. Most people know that both of these men were involved in politics as well as in the movie industry. For us to have this knowledge entails that the memory of each man is associated both to our thoughts and memories on politics as well as to those on movies. Nevertheless, thinking about Schwarzenegger is more likely to remind you of his movies than of his political career, whereas thinking about Reagan is more likely to remind you of his political career than of his acting career. Clearly, our memory trace for Schwarzenegger has a greater association strength to our movie memories than to our politic memories, whereas the opposite is true for the Reagan memory trace. We see something similar when we consider how thinking about IHOP is more likely to remind us of breakfast food than of burgers, whereas thinking about MacDonald's is more likely to remind us of burgers than of breakfast food. Of course, the thought of MacDonald's is still more likely to remind people of breakfast food than it is to remind them of the hula burger, even for people who have actually eaten the hula burger (yes, I will explain what a hula burger is in a minute). Part of the explanation is that most people's memory for hula burgers have incredibly weak trace strength, a concept that we will now elaborate on.

Section 5 (Minimum time = 25s)

The hula burger was a non-meat option that MacDonald's offered in order to reach out to Catholics, for Catholics abstain from meat on Friday's during the season of Lent. Hence, the Catholic could in good conscience order the hula burger, which had patties consisting of a mixture of cheese and pineapple. Because most people found this concoction rather unappetizing, it was quickly pulled from the menu, back in the 70's. Thus, people who have eaten it probably have not thought about it in 40 years. The memory has thus begun to fade. We have all experienced this with our school lessons as well—we might know our chemistry for the chemistry final, but have a hard time remembering that information once a year passes. The trace strength has weakened, but what exactly does this mean?

Section 6 (Minimum time = 40s)

One of the earlier ideas about what it is for a memory's trace strength to weaken was that the memory itself is decaying, thereby making it harder to "read" or to "see" the information contained in the memory. This conception was already present in Plato's writings (more than 2,000 years ago) and was still present during the dawn of the scientific study of psychology (a little over 100 years ago). However, a more useful analogy for our purposes is to think of each memory as having a light bulb hanging over it. These bulbs are special in that they have no 'off' switch, and are continuously turned on. If the light bulb is sufficiently bright, then you can read or see the content contained in the memory. However, with time, the bulb's battery will get used up and lose its 'trace strength,' causing the bulb to gradually get dimmer. Now, it seems (to pick back up our prior example) that as long as you are studying your chemistry, the light bulbs illuminating your chemistry memories do not get very dim. Why is this so? To better understand this, we will have to look deeper into the concept of <u>association strength</u>.

Section 7 (Minimum time = 50s)

I mentioned that thinking about MacDonald's is more likely to make you think of burgers than of, say, hash browns. But how come thinking about MacDonald's never causes you to think about thermo-nuclear fusion? Psychologists often think of it this way: When we process a memory (MacDonald's), we charge its trace strength. (Similarly, you charge your car battery whenever you run the engine). This incoming mental "electricity" does not stay contained in that cue memory, but spreads along links to associated memories. The incoming electricity will make the light bulb of an associated memory brighter, and possibly bright enough for us to read the content of that memory. Non-associated memories (like thermos-nuclear fusion) will not receive any of the electricity because they are not connected to the cue memory. Memories that are connected to the cue memory, but that have really dim bulbs, will require substantial processing of the cue in order get enough electricity for the memory to be recalled. This partially—but only partially—explains why we have a hard time remembering chemistry facts when we have gone a year without studying chemistry. However, if you never stop studying chemistry, then you continuously recharge the bulbs of your chemistry memories. But as I said, decrease in trace strength is not the major cause of our forgetting.

Section 8 (Minimum time = 60s)

Memory researchers have traditionally argued that it is not so much the passage of time that causes us to forget, but rather the learning of new information. Whenever you access a memory, thereby charging it, there is a fixed amount of electricity that it spreads in a given interval of time. Say, hypothetically (and definitely metaphorically), that a cue memory spreads 100 units of electricity per second. Also say hypothetically that MacDonald's is only associated to 2 memories: burger and hash browns. The reason thinking about MacDonald's causes you to think of burgers first is that the memory of burgers receives, say, 80 of the 100 units of electricity spread each second. That is, for a memory to have a stronger association to a cue memory just means that it receives a greater proportion of the electricity that the cue memory is dispersing. Hence, the reason thinking about MacDonald's reminds you of burgers before it reminds you of hash browns is that the memory bulb for the burger is receiving more of the electricity being spread and so gets brighter more quickly. Yet association strength is only one factor in the remembering process. For example, if you just ate hash browns 30 minutes ago, and so the trace strength of hash browns is already high, then the memory bulb for hash browns will already be almost bright enough to read its memory contents, and so it will only have to receive very little electricity. Hence, in this case, MacDonald's might remind you of hash browns first. In any event, now that we have a clearer grasp of the mental machinery underlying the remembering process, we are in a position to explain the major source of forgetting: interference.

Section 9 (Minimum time = 55s)

As we discuss interference, let us switch back to the scenario in which we are trying to learn chemistry. Let us pretend that we have only studied the first 10 elements of chemistry. A major cue you would use to remind yourself of these elements is obviously "chemistry." If each of the elements is equally associated to "chemistry," then, when "chemistry" is being processed as a cue, it will dispense [100 units of electricity ÷ 10 elements =] 10 units of electricity to each of the element memories (per second). Now, what happens if you move on from studying the first 10 elements and are now studying elements 11-20? We will assume that each of the first 20 elements is now equally well associated to the cue "chemistry." Thus, the "chemistry" cue will now only dispense [100 units of electricity ÷ 20 elements =] 5 units of electricity to each of the first 10 element memories (per second). Hence, learning elements 11-20 caused the "chemistry" cue to spread less electricity to the memories of the first 10 elements, down from 10 units per second to 5 units per second. Learning elements 11-20 has now made it harder to remember the first 10 elements. This is how new learning can cause forgetting. It's like the memories associated to the cue are *competing* for electricity and are *competing* to be remembered. As the number of competitors increases, the less likely a particular memory is to win the competition and be recalled.

Section 10 (Minimum time = 35s)

Now, let us assume that, seeing that we have weakened our memories for those first 10 chemistry elements, we decide to restudy those first 10 elements (but not elements 11-20). This restudying will do 2 things. First, it will strengthen the trace strengths of the first 10 elements, which by itself will make them more easily recallable on a quiz. It also, though, will strengthen the association strengths between chemistry and each of the first 10 elements. When we last finished studying the second set of elements, 'chemistry' was sending 5 units of electricity per second to each of the 20 elements. Let us say that our further studying of the first 10 elements has so strengthened their associations to 'chemistry' that they now each receive 9 units per second whenever we think about chemistry. As a direct consequence, each of the elements 11-20 is now only receiving 1 unit per second, for the 'chemistry' cue can only disperse 100 units total per second. Hence it might be that we are now unable to recall elements 11-20, or that we have forgotten them.

Section 11 (Minimum time = 30s)

What we have laid out are some basic principles of memory and forgetting. Traditional theories of forgetting have assumed that forgetting is caused primarily by interference from competing memories, which are other memories associated to the cue being used. Differences in trace strength are thought to be more important for determining how quickly a memory is recalled, rather than whether or not it is recalled. One must keep in mind that we have explained these principles in simple scenarios (e.g., our memory of MacDonald's being only associated to two other memories). When we say that it is the learning of new information that causes forgetting of old information, we may seem to imply that it is impossible to have expert knowledge of chemistry, or of different bird species. But of course, there are those who have that expert knowledge. We will just say that the associations and links connecting an expert's memories are considerably more complex than the examples we have been going over.

Experiment 1: Text 2

Section 1 (Minimum time = 200s)

The traditional account of forgetting (that we just covered) assumed that forgetting was caused primarily by the interference produced by competing memories. A more recent theory of forgetting has built on top of this work by adding a new forgetting process: memory inhibition (or memory suppression). In so doing, this theory has radically altered conceptions about forgetting. Before we dive more into the details, let us first consider the inspiration for this new theory.

I have explained that, in traditional accounts of forgetting, the main reason that a cue becomes incapable of reminding us of a particular memory is that it has become more strongly associated to other competing memories. According to the inhibition theorists, the problem that had been ignored is not why we forget things, but rather why we remember things as well as we do. For example, whenever we think of fruit, most of us immediately think of apples, bananas, and grapes. Yet, if we are asked about the fruit popular in Hawaii, we have no problem remembering the pineapple. This is striking because our impulse to think of the other fruits (apples, bananas) is essentially a habit, and we all know how hard it is to break habits. In the realm of bodily movement, habit-driven responses are over-ridden by inhibitory neurons. For example, when we accidentally knock something off a counter, we reflexively and without thought try to catch it before it hits the ground. But what if you knock over a cactus? In that case, the brain must send inhibitory messages to your arm in order to stop it from trying to catch the cactus. Similarly, we might inhibit our habitual response to think of apples when trying to retrieve our pineapple memory. Hence, whenever there is some specific memory that we are trying to retrieve, inhibitory mechanisms are thought to come into play and to further the success of that retrieval attempt.

Section 3 (Minimum time = 55s)

It is clear how inhibitory mechanisms might be involved in our memory machinery, but how do they work? According to the theory, the inhibition holds down the trace strengths of competing memories to a low level. By keeping the trace strengths of competing memories low—by keeping their bulbs dark—the inhibitory mechanism prevents the habit–driven retrieval of those memories. You are thus more capable of retrieving the memory you are interested in because you are not being distracted by those competing memories. Thinking about the competing memories as distractors is key to understanding the way inhibition works. Just as you're better able to process the information in a book if there is less distraction (e.g., party music, a spouse nagging you about chores, a strange odor seeping from a person nearby), so too are you better able to access the information contained in a memory trace of interest if competing memories are not springing into consciousness. We all know the feeling when the inhibitory mechanism doesn't work, such as when you are trying to remember a computer password you just recently made, but are unable to get the old password out of your head.

Section 3 (Minimum time = 50s)

This new theory prompts the old question: why do we forget? We are said to have this inhibitory mechanism that reduces the interference caused by competing memories, and yet it was just this interference that was traditionally blamed for our forgetfulness. So, if the competing memories are no longer the major cause of forgetting, then what is? But of course, it is the inhibition itself that is the main cause. The reason you have forgotten an older password is that, when first learning a new password, you had to repeatedly inhibit that old password. This reduction of the inhibited memory's trace strength (dimming of the bulb) has made it hard for it to get enough electricity later when you want to retrieve it. There is some speculation that inhibition might even make it difficult to charge the memory bulb of an inhibited memory. This would make sense because low trace strength, by itself, has long been thought to not substantially cause forgetting. In any event, it seems that inhibition not only facilitates the act of remembering desired information, but also impairs (later) acts of remembering other information.

Section 4 (Minimum time = 70s)

Considering that inhibition causes forgetting, one might wonder if inhibition is really that advantageous. The inhibition theorists would argue that it is, and that it is outdated information (as old passwords normally are) that are most negatively affected by inhibition. Inhibition can thus be thought of as decluttering the mind of old memory "files." Keep in mind though that "decluttering" is merely a metaphor, and that the inhibited memory traces are not thought to be removed in any sense; rather, inhibited memories merely have their trace strength kept low.

We have so far described the inhibition theory as it was first presented. The theory has developed since then and has gotten more specific on a number of points. First, not all competing memories are inhibited. On the contrary, inhibition is only thought to be applied to strong competitors, to memories that are causing substantial distraction to your conscious attempt to remember some bit of information. For example, when you are trying to remember a lesser known president (say, John Tyler), it would be beneficial if you could prevent Bush, Obama, and Trump from getting stuck in your head. However, you probably do not have to worry about being distracted by your memory of Chester Arthur (the 21st president), and so there is no strong need to inhibit that memory. You only inhibit a memory if doing so significantly aids your effort to remember some particular information of interest. Inhibition that is more extensive might harm your general ability to remember useful information.

Section 5 (Minimum time = 60s)

Research on the inhibition theory has also produced results which suggest that a change was needed in the very conception of what a memory trace is. What were those results? It was found that a person's memory of spear became less accessible as a result from thinking of a loud thing beginning with "si." The person had earlier in the experiment read that sirens are loud, and that was the word that they were supposed to recall. Doing so made them less likely to think of spear when they were later asked to think of something sharp that begins with "sp." But why should spear be inhibited when you are trying to think of a particular loud thing? Spears are not loud, and most people do not in any way associate them with loudness. What the researchers concluded was that spear was not inhibited. Instead, what were inhibited were other loud things that participants studied in the beginning of the experiment, namely, cannon, grenade, and gun. However, because spear shares characteristics with these inhibited memories—they are all weapons—spear too was affected by the inhibition. This insight was what prompted the inhibition theorists to change their conception of what a memory trace is.

Section 6 (Minimum time = 80s)

A simple way of viewing memories is to think of them as being unitary, to think of each as being a single file completely independent of every other memory. Indeed, this was the conception we explicitly used above. The inhibition theorists began to claim that memories are in fact distributive. There might be a single memory trace for each elementary idea, such as red, solid, spherical. However, more complex memories, such as apple, are composites made up of those more elementary memory features. Importantly, there is only one trace for each elementary feature. There are not two different memory copies of the red trace, but rather a single trace that is shared by, say, your apple memory and your ketchup memory. If you inhibit apple while thinking of another fruit (such as banana), then the red trace will be inhibited, for it was one of the traces making up the apple memory. But because the red trace is part of the ketchup memory, that means that our ketchup memory will also be inhibited to some degree. However, other elementary traces that are part of ketchup but are not shared with the apple memory, such as liquid, will remain uninhibited. The ketchup trace will therefore only be partially inhibited. It should also be noted that the features that apple shares with banana—the fruit that was remembered—will not be inhibited. Otherwise, the inhibitory system would inhibit part of the very memory that the person is trying to retrieve! Hence, elementary features such as sweetness, fruit, etc. will not be inhibited in the apple trace, and it will only be partially inhibited.

Section 7 (Minimum time = 80s)

The final development in the new inhibition theory of forgetting was the recognition that inhibition mechanisms, like all mechanisms in the human body, do not function flawlessly. In fact, as the demand for inhibition increases, the probability of successful inhibition decreases. Very strong competitor memories can be thought of as sometimes being too strong to be subdued. When inhibition fails, then one can in fact expect the non-inhibited memory in question to actually be more accessible than it was prior to the attempt to inhibit it. Of course, that makes it seem that a subsequent attempt to inhibit that memory is even less likely to be successful, with another inhibition failure making that competitor even stronger, and thus it seems like we will never be able to inhibit this memory unless we wait a little while for its trace strength to naturally dissipate. However, recent data suggests that those subsequent attempts to inhibit the memory also gain in forcefulness, and that eventually the attempts at inhibition will succeed and start to make that competitor memory less accessible.

Such is the developing inhibition theory of forgetting. It should be remembered that the theory basically assumes everything that the traditional interference theory of forgetting did, and then adds new mechanisms and ideas about how the memory machinery works. However, this new theory is radically different from the old theory insofar as it suggests that interference is only a minor cause of our forgetting, and that the major cause is inhibition. The older, more traditional theory of forgetting based solely on interference still has its defenders to this day, and the debate between the two camps does not seem to be ending any time soon.

Appendix B

Experiment 1: Text 1 Test

Question 1 (Minimum time: 80s)

George Willis is a philosophy professor at Tufts University. When he was a graduate student, he studied the works of Franz Brentano and of his pupils, including Tomas Masaryk. At the time, Willis found Tomas Masaryk to be an interesting character in world history, for Masaryk succeeded in gaining Czechoslovakian independence as a republic after World War I. Masaryk also both founded and was the first President of Czechoslovakia. In any event, George Willis has not studied or thought about Tomas Masaryk since his graduate school days, but has rather studied the works of various other philosophers, mostly those centered in Latin America. As a result, when George Willis is asked to name as many philosophers as he can, he always seems to forget Masaryk.

Will this learning of other philosophers also cause George Willis to forget to name Masaryk when asked to name as many Czech politicians as he can?

a. Yes, for those other philosophers have been more thoroughly studied, have stronger associations, and so receive too much of the mental electricity being spread

b. Yes, for too much time has passed and so the trace strength of Masaryk will be too low

c. No, for the learning of new information does not decrease the trace strength of old information

d. No, for much of Willis' subsequent learning is associated to his memories on philosophy, but not to his thought on Czechoslovakia

Question 2 (Minimum time: 80s)

Two students are studying for their second chemistry test. The first chemistry test covered their memory for the first 20 elements; their current test will cover their memory for elements 21-40. The first student has elements 21-40 written down on a sheet of paper and studies by just rereading that sheet of paper. The second student has made notecards, with one side asking a question (for example, “What is the 23rd element?”) and the other side providing the answer (Vanadium). The second student tests himself using these notecards, not even looking at the right answer if he is confident that he has answered the question correctly. Which of the following is correct?

a. The first student is likely to forget more of the first 20 elements than is the second student, for the first student is actually physically perceiving information about the second 20 elements, thereby strengthening their association to “Chemistry” for him

b. The second student is likely to forget more of the first 20 elements than is the second student, for the first student actually has to suppress those memories when trying to test himself over elements 21-40

c. Whenever a cue (such as “Chemistry Element” and “23”) leads to retrieval of information (Vanadium), that causes the association between the cue and that information to strengthen. Hence, both students are likely to forget much of the first 20 elements as a consequent of studying elements 21-40.

d. The form of study will not matter. What is important is that any study method for studying elements 21-40 means that the student is not studying the first 20 elements, and so their trace strengths consequently weaken.

Question 3 (Minimum time: 70s)

Aristotle reported that, as a teenager, he decided to tackle the Greek classic, Herodotus’ The Histories. This book is massive, weighing in at 736 pages. Aristotle thus decided to just read 10 pages a day, and so finished it within 2-3 months. Now, in this book one comes across many characters (think of our own history books), some prominent and some minor. Hence, one is likely to forget some characters from the middle of the book (covering the history of Scythia) as a consequence of reading about the characters from the remainder of the book (largely covering the Greaco-Persian wars). Which of the following is correct?

a. One's memory for prominent characters from the middle of the book will in not be affected from reading about characters from the remainder of the book, for they are prominent and memorable characters

b. Characters from the middle of the book, both minor and prominent, will become less accessible because they will receive less of the mental electricity

c. One's memory for the minor characters from the middle of the book will not be affected by interference from characters learned later because they were not that memorable to start with

d. Memory for the characters from the middle of the book will only become less accessible if you test yourself over the later characters.

Question 4 (Minimum time: 65s)

Two students are studying the various species of snakes for an upcoming exam. The test requires the students to give the names of 50 different snakes and to give their average lengths. One student studies an alphabetical list of these 50 snakes and their lengths. The other student divides the snakes into 5 groups, with each group representing the continent of origin. Hence, this student not only studies these snakes' names and lengths, but also whether they came from North America, South America, Africa, Europe, or Asia. Which of the following is correct?

a. The first student will perform better on the test because he is not wasting his time studying continents of origin

b. The second student will perform better because he can use the continents of origin as cues

c. The first student will perform better on the test because he is learning the snakes in a specific order

d. The second student will perform better because he is trying to encode more detailed memory traces

Question 5 (Minimum time: 65s)

A student of East-Asian religions has decided to learn Sanskrit so that he can read many of their sacred scriptures in their original language. He studies 10 Sanskrit words on one day and another 10 Sanskrit words the next day. On the third day, he tries to recall all 20 Sanskrit words. Which of the following is correct?

a. The cue "Sanskrit words" will send equal amounts of activation to words from the first day and to words from the second day

b. The cue "Sanskrit words" will send greater amounts of activation to words from the second day than to words from the first day

c. The cue "Sanskrit words" did not change in its association strength to the first set of words after the second set of words were studied; it is just that further associations were added to that cue

d. The words from the first day have been studied just as much as the words from the second, so there will be no difference in trace strength between those two groups of words

Experiment 1: Text 2 Test

Question 1 (Minimum time: 55s)

In a game of trivia, a person tries to remember the two actors starring in "The Gladiator." After correctly recalling Russell Crowe, the person is unable to remember Joaquin Phoenix. The person thus starts trying to think of other movies that Phoenix had been in, hoping that will help to jog his memory. Will the fact that the person just remembered Russell Crowe impair the ability of the other cues—these other movies that Phoenix starred in—to remind the person of Phoenix's name?

a. No, according to both the inhibition theory of forgetting and the traditional interference theory of forgetting

b. No, according to the inhibition theory, **but** yes, according to the traditional interference theory of forgetting

c. **Yes, according to the inhibition theory, but no, according to the traditional interference theory of forgetting**

d. Yes, according to both the inhibition theory of forgetting and the traditional interference theory of forgetting

Question 2 (Minimum time: 65s)

Timmy, an 8-year old boy has just seen the movie 101 Dalmatians for the first time, and loved it. He is now trying to memorize the names of all 101 Dalmatians, and has so far practiced encoding the first 20. Timmy's father, a memory researcher, is wondering what would happen if he made a computer program to help Timmy study. The program shows Timmy the names of 10 of the Dalmatians, showing one at a time for 5 seconds. Will this extra studying negatively affect Timmy's memory for the 10 Dalmatians not presented on the program, but which Timmy had already studied earlier?

a. No, according to both the inhibition theory of forgetting and the traditional interference theory of forgetting

b. No, according to the inhibition theory, but yes, according to the traditional interference theory of forgetting

c. Yes, according to the inhibition theory, **but** no, according to the traditional interference theory of forgetting

d. Yes, according to both the inhibition theory of forgetting and the traditional interference theory of forgetting

Question 3 (Minimum time: 120s)

A memory researcher ran an experiment. In this experiment, participants were presented with 36 category-example pairs (e.g., INSECT-beetle). Six different categories (e.g., WEAPON, TOOL, INSECT would be 3 of these 6 categories) were each presented 6 times, and were paired with a different example each of those 6 times (e.g., beetle, hornet, mosquito could be 3 of the 6 examples paired with INSECT, whereas, wrench, hammer, and nails could be 3 of the 6 examples paired with TOOL). Participants were given 5 seconds to study each of these 36 category-example pairs.

Importantly, half of the participants only studied examples that were weakly associated with the categories (e.g., INSECT-locust, INSECT-cicada, INSECT-tick) and half of the participants only studied examples that were strongly associated with the categories (e.g., INSECT-beetle, INSECT-hornet, INSECT-mosquito).

Next, participants were tested on half of the examples of each category. On each trial, the category and first letter of the appropriate example were presented as cues (e.g., "INSECT-c___?"). No corrective feedback was given. The test results did suggest that the researcher did a good job of choosing his examples: Those who studied the strong examples correctly responded to 95% of the test questions, whereas those who studied the weak examples only responded correctly to 60% of their test questions.

Finally, the researcher tested the participants over the other half of the examples of each category, again providing the category and the first letter of the appropriate examples on each of the test questions (e.g., INSECT—l__?). An interesting question is whether the participants' ability to recall these examples was impaired as a result of recalling the other examples on the first test. More specifically, did the participants who studied the strong examples receive a greater degree of such impairment than did the participants who studied the weak examples?

a. No, according to both the inhibition theory of forgetting and the traditional interference theory of forgetting

b. No according to the inhibition theory, **but** yes according to the traditional interference theory of forgetting

c. Yes according to the inhibition theory, **but** no according to the traditional interference theory of forgetting

d. Yes, according to both the inhibition theory of forgetting and the traditional interference theory of forgetting

Trick Question (Minimum time: 25s)

In a remote association task (RAT), participants are presented with 3 cues (e.g., "manners"-"tennis"-"round") and are asked to come up with a word that is associated to all 3 of the cues (e.g., "table"). This task is made more difficult when, prior to seeing the 3 cues, participants are exposed to distraction-words that are each associated with some but not all of the cues (e.g., "polite," "racket," "circle"). Pick the third answer. According to the inhibitory account, should people who exhibit substantial retrieval-induced forgetting be more negatively affected by the distraction words (than people who exhibit less substantial retrieval-induced forgetting), less negatively affected, or equally affected?

a) More negatively affected

b) Less negatively affected

c) Equally negatively affected

d) Cannot Tell

Question 4 (Minimum time: 85s)

Participants are asked to learn a number of "facts" about certain fictional characters. For example, they could study the following facts about the "salesman": the salesman sells apples; the salesman sells cigarettes; the salesman sells staplers; etc. They could also learn facts about the librarian: the librarian likes dressers; the librarian likes trees; the librarian likes skateboards; etc.

In the researchers' first experiment, participants learned 6 facts about each of 8 characters, and then took a test over half of the facts of each character (e.g., the salesman sells ci___?). However, because these materials do not have natural associations (as contrasted with, for example, the association everyone has between Fruit and Apple), the researchers were worried that the participants couldn't learn the material well. Hence, the researchers thought that it would be a good idea to run a second experiment at the same time, which was identical except participants only had to learn facts about 4 characters instead of 8 characters (they did increase the number of facts learned about each character to 10 in order somewhat maintain the total number of facts that had to be learned). Again, the participants took a test over half the facts of each character. According to the inhibition theory, what should we predict?

a. The untested facts will be inhibited to a greater degree in Experiment 1 than in Experiment 2.

b. The untested facts will be inhibited to a greater degree in Experiment 2 than in Experiment 1.

c. The untested facts will be inhibited to the same degree in both experiments

d. The principles of the inhibition theory do not make any predictions on this issue

Question 5 (Minimum time: 160s)

A memory researcher ran an experiment. In this experiment, participants were presented with 36 category-example pairs (e.g., COTTON-curtains), which the participants tried to memorize for a later test. Eight different categories (e.g., COTTON, LEATHER, LOUD, SHARP were 4 of these 8 categories) were each presented 6 times, and were paired with a different example each of those 6 times (e.g., curtains, pajamas, and robe were 3 of the 6 examples paired with COTTON, whereas, briefcase, saddle, and belt were 3 of the 6 examples paired with LEATHER; see the table below for 4 of the categories). Participants were given 5 seconds to study each of these 36 category-example pairs.

Cotton	Leather	Loud	Sharp
curtains	briefcase	jackhammer	needle
napkin	saddle	siren	tack
sheet	whip	traffic	thorn
(clothing[a])		(weapon[a])	
pajamas **(PT)**	belt	cannon **(PT)**	dagger
robe **(PT)**	boots	grenade **(PT)**	spear
slacks **(PT)**	skirt	gun **(PT)**	sword

Importantly, 3 examples from each category happened to share a category with 3 examples from another category. For example, *pajamas* (a COTTON example) and *belt* (a LEATHER example) both belong to the shared category of CLOTHING. These shared categories (e.g., CLOTHING, WEAPON) were never explicitly mentioned to the participants in the experiment.

The reason why the researcher used examples that shared these categories was that the researcher was interested in testing predictions of the traditional interference theory of forgetting. In particular, he wanted to test predictions that this theory makes when memories are assumed to have distributed representations as opposed to unitary representations.

After studying the category-example pairs, participants took a practice test over half of the examples of half the categories (e.g., 'COTTON-pa___?'). The figure above notes which examples were assessed during the practice test (PT; e.g., *pajamas, cannon*) and which were not (e.g., *curtains, belt*).

After the practice test, the participants took a final test over the other examples. In order to increase the difficulty of the test, cues different from the original categories were used (e.g., "HANGS-c___?" for *curtains* and "LOOP-b___?" for *belt*). How is performance on this test going to be affected from having worked through the practice test, according to the two prominent theories of forgetting? Remember that we are assuming that memories have a distributed representation rather than a unitary representation.

a. **<u>The inhibition theory</u>: belt has become more recallable than prior; <u>the traditional interference theory</u>: curtains has become more recallable than prior**

b. <u>The inhibition theory</u>: curtains has become **more** recallable than prior; <u>the traditional interference theory</u>: belt has become **less** recallable than prior

c. <u>The inhibition theory</u>: curtains has become **less** recallable than prior; <u>the traditional interference theory</u>: belt has become **more** recallable than prior

d. <u>The inhibition theory</u>: belt has become **less** recallable than prior; <u>the traditional interference theory</u>: curtains has become **less** recallable than prior

Appendix C

Experiments 2-4: Text 1

Section 1 (Minimum time = 10s)

We all are familiar with the act of remembering. In fact, most of us are annoyed by the limits of memory. However, scientists are discovering that forgetting might be adaptive. These scientists are of course also learning that our memory system is more complicated than you would think. This article will cover some of the insights made by these scientists.

Section 2 (Minimum time = 10s)

Memory scientists take certain things for granted. For example, it is assumed that you can only remember something if you were reminded of it by some other thing. When you remember that you have been meaning to call your mother, you do not just remember this out of the blue. Rather, you were thinking about something else which reminded you to call your mother. Seeing a can of Red Bull might remind you to call your mother if she drinks 5 Red Bulls every day. The thing that reminds you of something else is called a cue.

Section 3 (Minimum time = 20s)

It should be noted that a **cue** must already have a **link** to the memory it reminds you of. Thinking about monks is unlikely to make you think about the "macarena dance." (This assumes that you have not seen monks doing the macarena. And the "macarena" is not to be confused with the "monkarena," which monks do perform). Most of us have not made a mental connection (or **link**) between monks and the macarena. However, thinking about the cookies you're eating can clearly remind you of milk. Similarly, thinking about mice might cause you to think about cats. And mugs make us think more of beer or coffee than of tea or lemonade.

Now, if I ask you what comes to mind when you hear the word fire, there are many answers you could give. The word fire could remind you of camping. It could also remind you of Cheetos (the "flaming" hot ones). Or you might think of Hell. But, we can only remember one thing at a time. So what will you be reminded of first whenever you think about fire?

Section 4 (Minimum time = 30s)

Scientists have traditionally assumed that 2 factors determine which memory a **cue** reminds you of first. These are **link strength** and **trace strength**. "Trace" is here being used in the sense of a "memory trace." A memory trace is a record in one's mind for a particular memory. For example, we say that we store information about muzzles in one memory trace—the muzzle memory trace. Information about muggles is stored in a different memory trace—the muggle trace.

The concepts of **link strength** and **trace strength** can be illustrated by thinking about MacDonald's and IHOP. Both of these restaurants serve burgers as well as eggs. Importantly, most of us know (or can remember) this information about these two restaurants. Thus, our memory of each restaurant is **linked** to both our burger memories and to our egg memories.

Nevertheless, thinking about IHOP is more likely to remind us of eggs than of burgers. MacDonald's, however, is more likely to remind us of burgers than of eggs. Clearly, the **link** between IHOP and eggs has more **link strength** than does the **link** between IHOP and burgers. It is the opposite, though, with MacDonald's.

Put simply, IHOP has a stronger association to eggs than to burgers. MacDonald's has a stronger association to burgers than to eggs. We will flesh out the mechanism underlying **link**

strength later on.   But first: Why might MacDonald's sometimes makes us think of eggs more quickly than burgers?

Section 5 (Minimum time = 20s)

The reason why MacDonald's might remind us of its eggs first is that the eggs memory trace might have high **trace strength**. For example, perhaps you just got up and are hungry for breakfast. Then your mother asks you if you want something from MacDonald's. You immediately think of the fact that they serve eggs. (And then you yell at your mother to go get eggs). The reason, as mentioned, is that your prior thinking caused your eggs memory to have high **trace strength**. We will now see what that means.

Imagine that each of your memories has a light bulb hanging over it. Furthermore, there is no way to turn the bulb off. They are continuously turned on. If a given bulb is bright enough, then you can "read" the contents of its memory. However, with time, the bulb's battery will get used up. That is, the bulb (or memory) will lose its **trace strength**. The bulb will therefore get dimmer, making it harder to read the contents of its memory. This is partially why we barely remember, say, what we learned in our chemistry class. It seems like we would not forget out chemistry if we never stopped studying it. It seems like the light bulbs would not then get dimmer. This suggests that there are ways to recharge these bulbs. Indeed, there at least 2 ways to recharge a memory's **trace strength**.

Section 6 (Minimum time = 10s)

First, whenever we process a memory, we recharge its **trace strength**. Each time you restudy Helium, you recharge its bulb. (Similarly, you recharge your car battery whenever you run the engine). This incoming “mental electricity” can only recharge the bulb to a certain point: until it is fully charged. This happens rather quickly. After that, the mental electricity spreads down the **links** connected to that memory. In this way, other memories **linked** to the Helium memory are also recharged. This is the second way in which **trace strength** is recharged.

Section 7 (Minimum time = 10s)

As mentioned, you can only recall a memory if it has high **trace strength**. Only then can you read the contents of the memory. Memories with dimmer bulbs take longer to recharge. Thus, to recall these memories, you must process a **linked cue** memory for a longer time. And so, consider again whether fire reminds me first of Cheetos or of camping. Assume that I ate ten bags of Cheetos yesterday. Fire is thus likely to remind me of Cheetos first. This is how **trace strength** influences the memory-retrieval process. We will now explore how **link strength** also influences this process. In fact, we will see why **link strength** and NOT **trace strength** is the main cause of forgetting.

Section 8 (Minimum time = 11s)

Whenever you charge a **cue** memory, there is a limited amount of electricity it can send to **linked** memories. Say the **cue** memory spreads 100 units of mental electricity per second. We said that IHOP (**cue**) has greater **link strength** to eggs than to burgers. What this means is that this **cue** sends more electricity per second to eggs than to burgers. It might send 70 units per second to one and 30 units per second to the other. Hence, the eggs bulb (**trace strength**) will get brighter more quickly. We are now in a position to understand why we forget.

Section 9 (Minimum time = 20s)

The basic idea is that forgetting is caused by new learning. ‘Learning’ means the adding or strengthening of other **links** to some **cue**. Say we are learning chemistry for the first time. Our teacher initially has us learn the first 10 elements. A major **cue** you would use to remind yourself of these elements would be ‘chemistry.’ Assume that, after studying, all elements have the same **link strength** to ‘chemistry.’ Also recall that a **cue** can only spread a limited amount of electricity: 100 units per second. Thus, ‘chemistry’ would spread [100 ÷ 10 elements =] 10 units per second to each of the elements. Each has a **link strength** of 10 units per second. You then study elements 11-20. Eventually, all of the first 20 elements have the same **link strength** to ‘chemistry.’ Specifically, each has a link strength of [100 ÷ 20 elements =] 5 units per second to each of the elements. Thus, the **link strength** of the first 10 elements has gotten worse. The reason is that you **linked** 10 new elements to the **cue**.

Section 10 (Minimum time = 30s)

Something similar happens if you then only study elements 11-20 even more. This additional study will increase their **link strength**. Assume these elements now receive 9 units per second. That would mean that each of the first 10 elements now only receive 1 unit per second. Their **link strength** has again gotten worse, and again because of further memorizing of other information.

We can think of the memories **linked** to a **cue** as competing for electricity. More importantly, we can think of them as competing to be remembered. Memories with stronger **links** tend to be recalled more quickly. In our example, the student would tend to think elements 11-20 more quickly than elements 1-10. But if this is the case, then it will be hard to think of elements 1-10 at all. The reason is that you would first have to get elements 11-20 out of your head. Think about how hard it can be to remember a song when you have another song already stuck in your head. Furthermore, it is hard to get elements 11-20 out of your head if the **cue** you are using ('chemistry') keeps reminding you of them. Thus, you can be incapable of retrieving a memory if other memories are "getting in the way." Scientists think of these competing memories as causing interference. Hence, this theory of forgetting is called the **interference theory**.

Experiments 2-4: Text 2

Section 1 (Minimum time = 10s)

The traditional account of forgetting (that we just covered) assumed that forgetting is caused by **interference** from competing memories. A more recent theory of forgetting has built on top of this work. What it adds is a new forgetting process: memory **inhibition** (or memory suppression). In so doing, this theory has radically altered conceptions about forgetting. Let us begin by considering the inspiration for this new theory.

Section 2 (Minimum time = 20s)

In the **interference theory** of forgetting, a cue becomes incapable of reminding us of a particular memory when it becomes more strongly linked to other competing memories. This theory only explains why we forget. But why do we remember things as well as we do?

For example, U.S. citizens think of "cities," they normally think of U. S. cities. They may also think of a few foreign cities, such as Paris or London. However, the Egyptian city Cairo is not likely to come to mind. It is not likely to come to mind even when they think about Egypt. Rather, they will usually think of deserts, pyramids, Moses, and mummies. It is not our mental tendency (or our habit) to recall Cairo when thinking about Egypt or cities. Our tendency or habit is to recall those other memories. Yet, if we are asked to name Egyptian cities, most people, effortlessly, recall Cairo. This is striking because normally it is hard to break habits.

Section 3 (Minimum time = 10s)

In the realm of bodily movement, habit-driven responses are over-ridden by inhibitory neurons. For example, when we accidentally knock something off a counter, we reflexively try to catch it. But what if you knock over a cactus? Then, the brain must send inhibitory messages to your arm in order to stop it from trying to catch the cactus.   Similarly, we might inhibit our habitual response to think of Moses or Paris when trying to retrieve our Cairo memory. Hence, inhibition can aid our attempt to recall a given memory.

It is clear how inhibition might be involved, but how does it work?

Section 4 (Minimum time = 20s)

According to the theory, **inhibition** holds down the trace strengths of competing memories to a low level. **Inhibition** thereby prevents the habitual retrieval of those memories. You are thus more capable of retrieving the memory you are interested in.

Thinking about the competing memories as *distractors* is key. You're better able to process the content in a book if there is less distraction. So too are you better able to read the content in a memory if competing memories are not springing to mind.

We all know the feeling when the **inhibition** doesn't work. Think about trying to remember a computer password when you can't get the old password out of your head.

Section 5 (Minimum time = 30s)

This new theory prompts the old question. Why do we forget? **Inhibition** reduces the **interference** caused by competing memories. Yet, it was **interference** that was thought to cause forgetting. If competing memories are not a major cause of forgetting, then what is?

But of course, it is **inhibition** itself that is the main cause. Let us say you have successfully remembered a new computer password. Well, you inhibit the old password each time you remember the new one. The reduction in the that memory's trace strength has made it hard for it to get enough electricity later when/if you want to retrieve it.

It is also believed that **inhibition** makes it difficult for the memory bulb to even be charged by incoming mental electricity. That is, it is hard for its trace strength to even be increased. This makes sense because low trace strength, by itself, has been thought to not truly cause forgetting. In any event, **inhibition** does not only help us remember (right now). It also impairs (later) acts of remembering.

Section 6 (Minimum time = 10s)

Considering that inhibition causes forgetting, is it really that advantageous? The inhibition theorists argue that it is. They claim that that it is outdated information (as old passwords are) that is most negatively affected. Inhibition can thus be thought of as decluttering the mind of old memory "files." Keep in mind though that "decluttering" is merely a metaphor. Don't think of inhibited memory traces as being "removed." Rather, inhibited memories merely have their trace strength kept low.

Section 7 (Minimum time = 20s)

We have so far described the inhibition theory as it was first presented. The theory has been further developed since then.

First, not all competing memories are inhibited. Only memories causing substantial interference are inhibited.

For example, say you are trying to remember a lesser known president (say, John Tyler). It would be nice if you could prevent Bush, Obama, and Trump from getting stuck in your head. However, your memory of Chester Arthur (the 21st president) won't be very distracting. You therefore will not inhibit that memory.

You ONLY inhibit a memory IF doing so significantly aids your effort to recall a particular memory. Inhibition that is more extensive might harm your general ability to remember useful information.

Section 8 (Minimum time = 30s)

Research on the inhibition theory has also produced results which suggest that a change was needed in the very conception of what a memory trace is. What were those results?

It was found that a person's memory of spear became less accessible as a result from thinking of a loud thing beginning with "si." The person had earlier in the experiment read that sirens are loud, and that was the word that they were supposed to recall. Doing so made them less likely to think of spear when they were later asked to think of something sharp that begins with "sp."

But why should spear be inhibited when you are trying to think of a particular loud thing? Spears are not loud, and most people do not in any way associate them with loudness. What the researchers concluded was that spear was not inhibited.

Instead, what were inhibited were other loud things that participants studied in the beginning of the experiment, namely, cannon, grenade, and gun. However, because spear shares characteristics with these inhibited memories—they are all weapons—spear too was affected by the inhibition. This insight was what prompted the inhibition theorists to change their conception of what a memory trace is.

Recent data have also suggested that we change our conception of what a memory trace is. What were those data?

Section 9 (Minimum time = 30s)

It was found that a person's memory of spear became less accessible in an odd scenario. Specifically, the person recalled a "loud thing" beginning with "si." The person had earlier in the experiment read that sirens are loud, and that was the word he was supposed to recall. Doing so made him less likely to think of spear when later asked to think of something sharp that begins with "sp."

But why should spear be inhibited when you are trying to think of a particular "loud thing?" Spears are not loud. Most people do not in any way link them with loudness. What the researchers concluded was that spear was not inhibited.

Instead, what were inhibited were other loud things that participants studied. In particular, participants inhibited cannon, grenade, and gun. However, spear shares characteristics with these inhibited memories. They are all weapons. Because spear was similar to these other inhibited memories, it was negatively affected. This insight was what prompted the inhibition theorists to change their conception of what a memory trace is.

Section 10 (Minimum time = 20s)

A simple way of viewing memories is to think of them as being **unitary**. Each memory is thus a single file completely independent of every other memory. This was the conception I have been using. By contrast, memories can also be viewed as **distributive**.

In a **distributive** theory of memory, there might be a single memory trace for each elementary idea, such as *red*, *solid*, *spherical*. More complex memories, such as *apple*, are made up of elementary memory traces. That is, a complex memory is not a single trace, but a group of traces. Importantly, there is only one trace for each elementary idea. Consider your *apple* memory and your *ketchup* memory. Both contain the *red* trace. However, they do not each have their own unique *red* trace. Then you would have at least 2 different *red* traces. But we just said that each person only has one *red* trace. Hence, there must be a single *red* trace that is part of both memories.

Section 11 (Minimum time = 20s)

If you inhibit *apple* while recalling another fruit (e.g., *banana*), then the *red* trace will be inhibited. Inhibition of a complex memory is the inhibition of (most of) the elementary traces making it up. (We will discuss why some of the elementary traces are not inhibited). But the *red* trace is also part of the *ketchup* memory. Thus, our *ketchup* memory will also be inhibited to some degree. However, other elementary traces that are part of *ketchup* but are not shared with the *apple* memory (e.g., *liquid*) will remain uninhibited. The ketchup trace will thus be partially inhibited.

It should be noted that the features that apple shares with banana will not be inhibited. Otherwise, one would inhibit part of the very memory one is trying to retrieve! Hence, elementary features such as sweetness, solid, etc. will not be inhibited in the apple trace. It will only be partially inhibited.

Section 12 (Minimum time = 30s)

The final development in the theory was that inhibition is not flawless. Memories causing stronger interference are harder to inhibit. Strong competitors can sometimes be too strong to be subdued. When inhibition fails, the strong competitor will cause more interference than it did prior to the inhibition attempt.

But would that not make a subsequent attempt to inhibit that memory even harder? And if that attempt fails, then that competitor will be even stronger. It thus seems like we might never be able to inhibit this memory. However, recent data suggests that successive attempts to inhibit a memory gain in forcefulness. Eventually, the attempts at inhibition will succeed.

Such is the developing inhibition theory of forgetting. It should be remembered that the theory basically assumes everything that the traditional interference theory of forgetting did, and then adds new mechanisms and ideas about how the memory machinery works. However, this new theory is radically different from the old theory insofar as it suggests that interference is only a minor cause of our forgetting, and that the major cause is inhibition. The older, more traditional theory of forgetting based solely on interference still has its defenders to this day, and the debate between the two camps does not seem to be ending any time soon.

Appendix D

Experiments 2-4: Text 1 Argument Inference Test

Question 1 (Minimum time: 30s)

George Willis is a philosophy professor at Tufts University. When he was a graduate student, he studied the works of Franz Brentano and of his pupils, including Tomas Masaryk. At the time, Willis found Tomas Masaryk to be an interesting character in world history, for Masaryk succeeded in gaining Czechoslovakian independence as a republic after World War I. Masaryk also both founded and was the first President of Czechoslovakia. In any event, George Willis has not studied or thought about Tomas Masaryk since his graduate school days, but has rather studied the works of various other philosophers, mostly those centered in Latin America. As a result, when George Willis is asked to name as many philosophers as he can, he always seems to forget Masaryk.

Will this learning of other philosophers also cause George Willis to forget to name Masaryk when asked to name as many Czech politicians as he can?

a. Yes, for those Latin American philosophers have been more thoroughly studied, and so have stronger associations to the cue (Czech politicians), and so receive too much of the mental electricity being spread

b. No, for the learning of new information does not decrease the trace strength of old information

c. No, for much of Willis' subsequent learning is associated to his memories on philosophy, but not to his thought on Czechoslovakia

Question 2 (Minimum time: 30s)

Aristotle reported that, as a teenager, he decided to tackle the Greek classic, Herodotus' The Histories. This book is massive, weighing in at 736 pages. Aristotle thus decided to just read 10 pages a day, and so finished it within 2-3 months. Now, in this book one comes across many characters (think of our own history books), some prominent and some minor. Hence, one is likely to forget some characters from the middle of the book (covering the history of Scythia) as a consequence of reading about the characters from the remainder of the book (largely covering the Greaco-Persian wars). Which of the following is correct?

a. One's memory for prominent characters from the middle of the book will in not be affected from reading about characters from the remainder of the book, for they are prominent and memorable characters

b. Characters from the middle of the book, both minor and prominent, will become less accessible because they will receive less of the mental electricity when trying to remember characters from the book

c. One's memory for the minor characters from the middle of the book will not be affected by interference from characters learned later because they were not that memorable to start with

Question 3 (Minimum time: 25s)

Two students are studying the various species of snakes for an upcoming exam. The test requires the students to give the names of 50 different snakes and to give their average lengths. One student studies an alphabetical list of these 50 snakes and their lengths. The other student divides the snakes into 5 groups, with each group representing the continent of origin. Hence, this student not only studies these snakes' names and lengths, but also whether they came from North America, South America, Africa, Europe, or Asia. Which of the following is correct?

a. The first student will perform better on the test because he is not wasting his time studying continents of origin

b. The second student will perform better because he can use the continents of origin as cues

c. The first student will perform better on the test because he is learning the snakes in a specific order

Question 4 (Minimum time: 30s)

Two students are studying for their second chemistry test. The first chemistry test covered their memory for the first 20 elements; their current test will cover their memory for elements 21-40. The first student has elements 21-40 written down on a sheet of paper and studies by just rereading that sheet of paper. The second student has made notecards, with one side asking a question (for example, “What is the 23rd element?”) and the other side providing the answer (Vanadium). The second student tests himself using these notecards, not even looking at the right answer if he is confident that he has answered the question correctly. Which of the following is correct?

a. The first student is likely to forget more of the first 20 elements than is the second student, for the first student is actually physically perceiving information about the second 20 elements, thereby strengthening their association to “Chemistry” for him

b. The second student is likely to forget more of the first 20 elements than is the second student, for the first student actually has to suppress those memories when trying to test himself over elements 21-40

c. Whenever a cue (such as “Chemistry Element” and “23”) leads to retrieval of information (Vanadium), that causes the association between the cue and that information to strengthen. Hence, both students are likely to forget much of the first 20 elements as a consequent of studying elements 21-40.

Question 5 (Minimum time: 30s)

A student of East-Asian religions has decided to learn Sanskrit so that he can read many of their sacred scriptures in their original language. He studied 10 Sanskrit words on one day and another 10 Sanskrit words the next day. On the third day, he tries to recall all 20 Sanskrit words. Which of the following is correct?

a. The second set of 10 words will not be as easily remembered as they otherwise would be had the first set of 10 words not also been learned, for the first set is using up some of the mental electricity

b. The second set of 10 words will be more easily remembered than they otherwise would be due to the fact that another set of 10 words was learned first, for this earlier learning gave the student practice with the language

c. The ability to remember the second set of 10 words will in no way be affected by the earlier learning of the first set of words; for the learning of those words to cause interference, those words would had to have been learned ***after*** the second set of words was studied.

Experiment 3: Text 1 Memory Test

Question 1 (Minimum time: 30s)

Which of the following illustrates trace strength?

a. You are likely, when asked about your friend, to of the time when he had a job as a birthday party clown because that is a very odd job to take

b. You are likely to think of your mother when seeing a can of red bull because you know that your mother drinks a lot of red bull

c. You are likely to think of Cheetos when asked to name some foods because you recently ate a bunch of Cheetos

Question 2 (Minimum time: 30s)

Which of the following illustrates link strength?

a. When asked to list as many governors as you can, you are likely to think of Arnold Schwarzenegger if you just watched one of his movies

b. When visiting a catholic school, you wonder if they have nuns there who hit students' knuckles with rulers, because that is how you have seen catholic schools portrayed on TV

c. If you saw a movie with a really bizarre scene, you are likely to think of that scene when asked about the movie

Question 3 (Minimum time: 25s)

What is the main reason, according to the article, that we forget??

a. The trace strength of our memories continuously decrease unless we retrieve them from time to time

b. We strengthen links to competing memories and add new links to new competing memories

c. We repress the memories of things that we do not want to remember

Question 4 (Minimum time: 30s)

How does the trace strength of a memory get increased?

a. We focus out attention on the content of the memory

b. We focus our attention on a different memory that is connected to it by a link

c. Both of the above

Question 5 (Minimum time: 30s)

Say you study the first 10 elements of chemistry (i.e., elements 1-10) for an hour total. Three days later, you study the next 10 elements (i.e., elements 11-20), also for an hour total. Which of the following is most likely to be the case immediately after studying elements 11-20 for an hour?

a. The link strengths of the first 10 elements are equal to the link strength of the other 10 elements

b. The link strengths of the first 10 elements are greater than the link strength of the other 10 elements

c. The link strengths of the first 10 elements are weaker than the link strength of the other 10 elements

Experiment 4: Text 1 Non-Argument Inference Test

Note: The bolded content is the feedback given to the Ps *after* answering a question

Question 1 (Minimum time: 30s)

George Willis is a philosophy professor at Tufts University. When he was a graduate student, he studied the works of Franz Brentano and of his pupils, including Tomas Masaryk. At the time, Willis found Tomas Masaryk to be an interesting character in world history, for Masaryk succeeded in gaining Czechoslovakian independence as a republic after World War I. Masaryk also both founded and was the first President of Czechoslovakia. In any event, George Willis has not studied or thought about Tomas Masaryk since his graduate school days, but has rather studied the works of various other philosophers, mostly those centered in Latin America. As a result, when George Willis is asked to name as many philosophers as he can, he always seems to forget Masaryk.

Will this learning of other philosophers also cause George Willis to forget to name Masaryk when asked to name as many Czech politicians as he can?

a. Yes, that will make it harder to remember Masaryk

b. No, that will make it easier to remember Masaryk

c. No, that will have no impact on his memory of Masaryk, **for much of Willis' subsequent learning is associated to his memories on philosophy, but not to his thought on Czechoslovakia**

Question 2 (Minimum time: 30s)

Aristotle reported that, as a teenager, he decided to tackle the Greek classic, Herodotus' The Histories. This book is massive, weighing in at 736 pages. Aristotle thus decided to just read 10 pages a day, and so finished it within 2-3 months. Now, in this book one comes across many characters (think of our own history books), some prominent and some minor. Hence, one is likely to forget some characters from the middle of the book (covering the history of Scythia) as a consequence of reading about the characters from the remainder of the book (largely covering the Greaco-Persian wars). Which of the following is correct?

a. One's memory for prominent characters from the middle of the book will in not be affected from reading about characters from the remainder of the book

b. Characters from the middle of the book, both minor and prominent, will become less accessible, **because they will receive less of the mental electricity when trying to remember characters from the book**

c. One's memory for the minor characters from the middle of the book will not be affected by interference from characters learned later

Question 3 (Minimum time: 25s)

Two students are studying the various species of snakes for an upcoming exam. The test requires the students to give the names of 50 different snakes and to give their average lengths. One student studies an alphabetical list of these 50 snakes and their lengths. The other student divides the snakes into 5 groups, with each group representing the continent of origin. Hence, this student not only studies these snakes' names and lengths, but also whether they came from North America, South America, Africa, Europe, or Asia. Which of the following is correct?

a. The first student will perform better on the test

b. The second student will perform better **because he can use the continents of origin as cues**

c. Thee will be no difference between the performances of the two students

Question 4 (Minimum time: 30s)

Two students are studying for their second chemistry test. The first chemistry test covered their memory for the first 20 elements; their current test will cover their memory for elements 21-40. The first student has elements 21-40 written down on a sheet of paper and studies by just rereading that sheet of paper. The second student has made notecards, with one side asking a question (for example, "What is the 23rd element?") and the other side providing the answer (Vanadium). The second student tests himself using these notecards, not even looking at the right answer if he is confident that he has answered the question correctly. Which of the following is correct?

a. Only the first student is likely forget some of the first 20 elements

b. Only the second student is likely forget some of the first 20 elements

c. Whenever a cue (such as "Chemistry Element" and "23") leads to retrieval of information (Vanadium), that causes the association between the cue and that information to strengthen. Hence, both students are likely to forget much of the first 20 elements as a consequent of studying elements 21-40.

Question 5 (Minimum time: 30s)

A student of East-Asian religions has decided to learn Sanskrit so that he can read many of their sacred scriptures in their original language. He studied 10 Sanskrit words on one day and another 10 Sanskrit words the next day. On the third day, he tries to recall all 20 Sanskrit words. Which of the following is correct?

a. The second set of 10 words will not be as easily remembered as they otherwise would be had the first set of 10 words not also been learned, **for the first set is using up some of the mental electricity**

b. The second set of 10 words will be more easily remembered than they otherwise would be due to the fact that another set of 10 words was learned first

c. The ability to remember the second set of 10 words will in no way be affected by the earlier learning of the first set of words

Experiments 2-4: Final Test

Question 1 (Minimum time: 45s)

In a game of trivia, a teenager has to remember the main protagonist starring in “The Gladiator.” He correctly recalls that it was Russell Crowe that played the protagonist (and NOT Joaquin Phoenix, who is the other star actor in that movie).

The game swiftly moves on the next player's turn, and no mention is made of Phoenix.

After the game, the teenager walks back to his home. There, his parents ask him who starred in the movie "Walk the Line." The correct answer is Joaquin Phoenix, and the teenager does have an association between Phoenix and "Walk the Line" in his memory system. (Note also that Russell Crowe was NOT in this movie).

Will the fact that the teenager remembered Crowe's name during the game of trivia make it a little more difficult to correctly answer this new question?

a. No, according to the inhibition theory, **but** yes, according to the traditional interference theory of forgetting

b. **Yes, according to the inhibition theory, but no, according to the traditional interference theory of forgetting**

c. Yes, according to both the inhibition theory of forgetting and the traditional interference theory of forgetting

Question 2 (Minimum time: 40s)

Timmy, an 8-year old boy has just seen the movie 101 Dalmatians for the first time, and loved it. He is now trying to memorize the names of all 101 Dalmatians, and has so far practiced encoding the first 20. Timmy’s father, a memory researcher, is wondering what would happen if he made a computer program to help Timmy study. The program shows Timmy the names of 10 of the Dalmatians, showing one at a time for 2 seconds. This computer program cycles through the set of 10 Dalmatians 6 times. Timmy then tries to remember the 10 Dalmatians that he did not just study on the computer, obviously using Dalmatians as a cue. Will Timmy's ability to recall those Dalmatians be hindered by his engagement with the computer program?

Assume that there is no inhibition failure.

a. One's memory for prominent characters from the middle of the book will in not be affected from reading about characters from the remainder of the book

b. Characters from the middle of the book, both minor and prominent, will become less accessible, **because they will receive less of the mental electricity when trying to remember characters from the book**

c. One's memory for the minor characters from the middle of the book will not be affected by interference from characters learned later

Question 3 (Minimum time: 45s)

(i) Participants had to memorize a list of 10 insects.

(ii) Some participants studied the names of insects that are frequently encountered by us, and that are (for us) strongly associated to our idea of insects (e.g., beetle, ant, mosquito).

(iii) Other participants studied the names of insects that are much less common (e.g., cicada, tick, locust).

(iv) Next, every participant took a test in which they had to recall 5 of the 10 insects that were on his/her list. On each test question, the participants were given the first letter of a specific insect that they were being tested over (e.g., "INSECT-c___?" would be a test question over cicada). There were 5 of these questions; participants had to recall 5 specific insects from the list and NOT just any 5 from the list.

(v) Not surprisingly, participants who studied a list of common insects did much better on the test--they were each able to remember all 5 insects being tested. The other participants could only remember 2 of the insects that they were tested over; on three of the test questions, these participants did not remember a thing.

(vi) What was surprising, however, was that the participants who studied the common insects had their memory for the untested insects more negatively affected by taking the test than was the case for the participants who studied a list of less common insects. Is this second result to be expected? Assume that no inhibition failure occurs.

a. No according to the inhibition theory, **but** yes according to the traditional interference theory of forgetting

b. Yes according to the inhibition theory, **but** no according to the traditional interference theory of forgetting

c. Yes, according to both the inhibition theory of forgetting and the traditional interference theory of forgetting

Trick Question (Minimum time: 25s)

In a remote association task (RAT), participants are presented with 3 cues (e.g., “manners”-“tennis”-“round”) and are asked to come up with a word that is associated to all 3 of the cues (e.g., “table”). This task is made more difficult when, prior to seeing the 3 cues, participants are exposed to distraction-words that are each associated with some but not all of the cues (e.g., “polite,” “racket,” “circle”). This fourth question is a trick question, and I want you to give all twelve points to the third option. According to the inhibitory account, should people who exhibit substantial retrieval-induced forgetting be more negatively affected by the distraction words (than people who exhibit less substantial retrieval-induced forgetting), less negatively affected, or equally affected?

a) More negatively affected
b) Less negatively affected

c) Equally negatively affected

Question 5 (Minimum time: 50s)

Participants were asked to learn a number of “facts” about certain fictional characters. For example, they could study the following facts about the “salesman”: the salesman sells apples; the salesman sells cigarettes; the salesman sells staplers; etc. They could also learn facts about the librarian: the librarian likes dressers; the librarian likes trees; the librarian likes skateboards; etc.

In Experiment 1, participants learned 6 facts about each of 8 characters. In Experiment 2, participants learned 12 facts about each of 4 characters. Hence, in both experiments, participants had to learn 48 facts.

In both experiments, participants took a test over half the facts of each character, and the test determined which facts each participant would have to try to remember. For example, "the librarian likes dr___?" would test a participant to see if they remember that "the librarian likes dressers." Thus, participants from (say) Experiment 1 had to recall and report 3 specific facts about the librarian, and NOT just any 3 facts about the librarian.

Participants had their memory for the untested facts more negatively affected <u>from taking the test</u> if they were in Experiment 2 than if they were in Experiment 1. Is this predicted by the inhibition theory? Assume that no inhibition failure occurs.

a. No; there should be more inhibition in Experiment 1 than in Experiment 2

b. Yes; there should be more inhibition in Experiment 2 than in Experiment 1

c. No; the untested facts should be inhibited to the same degree in both experiments

Question 6 (Minimum time: 50s)

(i) Participants study a list consisting of COTTON things (e.g., "COTTON-curtains") and of LEATHER things (e.g., "LEATHER-suitcase"; see full list below).

(ii) Importantly, half of the COTTON items happen to also be instances of CLOTHING, as do half of the LEATHER items (see figure below).

(iii) Participants take a practice test (PT) over some of the items ("COTTON-p___?" for pajamas). Incidentally, the only items tested on the practice test are COTTON items that happen to be instances of CLOTHING (see figure below).

Cotton	Leather	Practice Test	Final Test
curtains	briefcase		
napkin	saddle		
sheet	whip	Cotton-pa___?	HANGS-c___?
(clothing)		Cotton-ro___?	LOOP-b___?
pajamas	belt		
robe	boots	Cotton-sl___?	
slacks	skirt		

(iv) After the practice test, the participants took a final test over the other examples. In order to increase the difficulty of the test, cues different from the original categories were used (e.g., “HANGS-c___?” for curtains and “LOOP-b___?” for belt).

(v) How is performance on this test going to be affected from having worked through the practice test, according to the two prominent theories of forgetting? Assume that memories have a distributed representation rather than a unitary representation only when considering the inhibition theory of forgetting.

a.) Inhibition Theory: curtain has been **unaffected**
Interference Theory: belt has become **less** recallable

b.) Inhibition Theory: curtain has become **less** recallable
Interference Theory: curtain has become **less** recallable

c.) Inhibition Theory: curtain has become more recallable
Interference Theory: curtain has been unaffected

Appendix E

Participants

Twenty-nine undergraduates (21 women) from the University of Oklahoma participated. They ranged from 18 to 21 years of age (M = 18.6). The Ps were predominantly White (79%), with the remaining Ps being American Indian or Native American (3.8%), Asian (11.5%), Black or African American (7.7%), and not reporting (3.8%). Ps received class credit for participating.

Design

A 2-group (Reading condition; Non-Reading condition) between-subjects design was used.

Procedure

Ps were randomly assigned to be either in the Reading condition or in the Non-Reading condition. Those in the Reading condition began by reading the two texts used in Experiments 2-4 (see Appendix C). They then took a test, which was the same as the final test used in Experiments 2-4 (see Appendix D). Ps in the Non-Reading condition, by contrast, began by taking the test, without getting to read the relevant texts first.

Results

Ps in the Reading group did not score significantly higher (M = 22.3, SD = 10.1) than did Ps in the Non-Reading group (M = 21.4, SD = 7.9), $t(28) = .25$, $p = .40$. This pattern was not changed when the data were reanalyzed after one P, who had a final test score of 0, was removed the data set. Again, Ps in the Reading group did not score significantly higher (M = 23.8, SD = 8.6) than did Ps in the Non-Reading group (M = 21.4, SD = 7.9), $t(27) = .74$, $p = .23$.