The Effect that Testing has on Nondeclarative Memory

David Smith
University of Maine, dsmit1728@gmail.com

Follow this and additional works at: https://digitalcommons.library.umaine.edu/etd

Part of the Cognitive Psychology Commons

Recommended Citation
Smith, David, "The Effect that Testing has on Nondeclarative Memory" (2020). Electronic Theses and Dissertations. 3305.
https://digitalcommons.library.umaine.edu/etd/3305

This Open-Access Thesis is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of DigitalCommons@UMaine. For more information, please contact um.librarytechnical.services@maine.edu.
THE EFFECT THAT TESTING HAS ON NONDECLARATIVE MEMORY

By

David B. Smith

B.S. Michigan State University, 2014

A DISSERTATION

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (in Psychology)

The Graduate School
The University of Maine
December 2020

Advisory Committee:

Shawn Ell, Ph.D., Associate Professor of Psychology, Advisor
Thane Fremouw, Ph.D., Associate Professor of Psychology, Chair
Jordan LaBouff, Ph.D., Associate Professor of Psychology
Mollie Rubem, Ph.D., Assistant Professor of Psychology
Benjamin Guenther, Ph.D
Testing has been shown to improve long-term memory retention by decreasing the amount of material forgotten, a phenomenon known as the testing effect. This positive impact of testing has been shown using direct tests of memory that require declarative memory, things like memorizing word-pairs and single-word lists. This dissertation is the first research to investigate how testing impacts nondeclarative memory using three experiments. The first and second experiment utilize the word fragment completion task to measure the effect that testing has on words learned via methodology thought to recruit either declarative or nondeclarative memory. The third experiment utilizes a probabilistic category learning task known as the weather prediction task. This task requires that participants learn the associations between cues and weather patterns and can be completed with either declarative or nondeclarative memory strategies. This research is the first of its kind investigating the role that testing has on nondeclarative memory. The results from Experiment 1 showed an overall benefit of testing relative to studying for both declarative and nondeclarative memory, but not a testing effect. Experiment 2 expanded upon these findings and assessed whether participants were using declarative or nondeclarative memory during the word fragment completion task. A testing effect was shown for the participants using declarative and nondeclarative memory, overall, there was less forgetting for the word fragments that were tested on as compared to ones studied.
Experiment 3 showed no decrease in performance over two days for those using nondeclarative memory regardless of if they learned from testing or studying. Taken together, these results show that testing can decrease forgetting for nondeclarative memory as well as declarative memory. This is important for multiple reasons, education being one of them. Learning in the classroom involves both declarative and nondeclarative memory, understanding the most effective way to improve nondeclarative memory has important implications for education. In addition, this research expands the large testing effect literature by expanding it to another memory system.
ACKNOWLEDGEMENTS

David would like to acknowledge the members of his dissertation committee: Mollie Ruben, Jordan LaBouff, Thane Fremouw, Benjamin Guenther, and his research advisor and mentor Shawn Ell. He would also like to acknowledge all the undergraduate research assistants who have helped make this research possible. Also, he would like to acknowledge his support system, Andrew Kaplan, Mark Schonfeld, Tyler Rhinesmith, Joshua Dennett, and Sparty. He would like to give a special thanks to Colin Bosma and Rachael Huff, whose support throughout graduate school has been instrumental for his success. Finally, he would like to acknowledge his mother, Lisa, father, Bryan, and sisters, Rachel and Caitlin. Their love and support have made all of this possible.
**TABLE OF CONTENTS**

ACKNOWLEDGEMENTS........................................................................................................... ii

LIST OF TABLES ..................................................................................................................... vii

LIST OF FIGURES ................................................................................................................... viii

CHAPTER 1 ................................................................................................................................. 1

INTRODUCTION ......................................................................................................................... 1

The Importance of Testing........................................................................................................... 2

Memory Systems ......................................................................................................................... 6

Evidence for a Dissociation.......................................................................................................... 9

Testing Effect Methodologies .................................................................................................. 20

Testing Effect Theories ............................................................................................................ 30

Summary .................................................................................................................................. 38

Transfer Appropriate Processing .............................................................................................. 39

Desirable Difficulties .................................................................................................................. 39

Distribution-based Bifurcation Model ......................................................................................... 39

Dual Systems Theory ................................................................................................................ 39

Conclusion ................................................................................................................................. 40

EXPERIMENT 1 ......................................................................................................................... 41

Word Fragment Completion Task ............................................................................................. 41

Method ...................................................................................................................................... 44
Memory Systems.............................................................................................................................................. 105
Limitations......................................................................................................................................................... 107
Future Directions ............................................................................................................................................. 109
General Experiment Conclusions .................................................................................................................. 112
References ....................................................................................................................................................... 115
Appendix ......................................................................................................................................................... 127
  A. The word list used in Experiment 1, adapted from (Roediger et al., 1992).................. 127
  B. Strategy Analysis ....................................................................................................................................... 129
Biography of the Author...................................................................................................................................... 130
LIST OF TABLES

Table 1. Brief descriptions of different theories that attempt to explain the testing effect ........39

Table 2: Number of participants in each of the conditions, as recruited from either Amazon’s
Mechanical Turk or the University of Maine. .................................................................54

Table 4. Total frequency of each pattern, the number of times each pattern occurred, and the
outcomes for Experiment 2 (Gluck et al., 2002)..................................................................74

Table 5: Descriptive statistics for Experiment 2.....................................................................79

Table 6: Descriptive statistics for participants fit by either a nondeclarative, declarative, or
guessing model. Included are mean accuracy and the percent of participants who adopted the
strategy. .................................................................................................................................82
LIST OF FIGURES

Figure 1: Results from Experiment 1 (Roediger & Karpicke, 2006a) ........................................3
Figure 2: Depiction of the task used in Experiment 1 of Poldrack and colleagues, 2001 ..........12
Figure 3: Stimuli used in (Reber et al., 2003) .........................................................................14
Figure 4: The distribution-based bifurcation model (Halamish & Bjork, 2011; Kornell et al., 2011) ...........................................................................................................................................34
Figure 5: Procedure for both the immediate condition (top) and the delay condition (bottom)....43
Figure 6: Performance for both the studied and tested words during the final evaluation phase for the declarative memory condition ........................................................................................................48
Figure 7: Performance for both the studied and tested words during the final evaluation phase for the declarative memory condition ........................................................................................................50
Figure 8: Procedure for both the immediate condition (top) and the delay condition (bottom)....57
Figure 9: Performance for both the studied and tested words during the final evaluation phase for the declarative memory condition ........................................................................................................60
Figure 10: Performance for both the studied and tested words during the final evaluation phase for the nondeclarative memory condition ........................................................................................................62
Figure 11: Performance for both the studied and tested words during the final evaluation phase for the test aware participants in the nondeclarative memory condition .............................................65
Figure 12: Performance for both the studied and tested words during the final evaluation phase for the test unaware participants in the nondeclarative memory condition ......................................66
Figure 13: Weather patterns used in Experiment 3 .......................................................................72
Figure 14: Geometric cues used in Experiment 3 .........................................................................72
Figure 15: Percent Optimal during the testing phase, split by timing and test conditions ..........80
Figure 16: Percent Optimal during the testing phase, split by timing and test conditions, for
those using a nondeclarative strategy......................................................................................................................83
CHAPTER 1

INTRODUCTION

Consider how a student learns material in a typical classroom setting. The student is taught the material, studies it outside of the classroom, and is then tested on it at a later date. While this is a common way for students to prepare for a final exam, it is not optimal. Instead, the student would be better off by preparing using practice tests or by quizzing themselves using flash cards. A well-researched phenomenon, known as the testing effect, refers to decreased forgetting after successful retrieval (testing) and has been the target of a large number of studies (Roediger & Karpicke, 2006b).

Studies investigating the testing effect have been limited in both the types of materials used and the memory systems recruited. These studies have used tasks thought to require declarative memory, the memory for information that can be intentionally and consciously recollected (Cohen & Squire, 1980). It is unclear the effect that testing has on nondeclarative memory, the knowledge that does not require conscious recollection and is marked by changes in performance due to experience (Packard, Hirsh, & White, 1989).

The primary goal of this dissertation is to fill the gap in the literature by investigating the testing effect with nondeclarative memory tasks and answer the following research questions: Does testing reduce the amount of forgetting relative to studying with nondeclarative memory tasks? Is long term retention greatest after testing (testing is a process in which one’s knowledge or skills are assessed) relative to studying (studying refers to the act of reengaging with material or a process that has already been learned)? Does the type of learning strategy affect memory retention after testing?
The Importance of Testing

Recently, there has been an increased reliance on testing to make both high- and low-stakes decisions. Consider the ramifications of the General Record Examinations (GRE) or the Medical College Admission Test (MCAT). These tests play a largely influential role as determinates for college and graduate schools’ admissions. How an individual performs on these tests can dictate not only which school they are admitted to, but if they are admitted at all. The No Child Left Behind act of 2001 and Common Core standards have led to controversy and to the misuse of tests in high-school classroom settings (Coburn, Hill, & Spillane, 2016; Marsh, Roediger, Bjork, & Bjork, 2007). Using testing as only a measure of performance in high-stakes decision making areas discounts the benefits of using testing as a learning tool.

One way that testing can be important for learning is that the actual act of taking a test can cause a greater retention of materials as compared to spending the equivalent amount of time re-reading and re-studying materials (Roediger & Karpicke, 2006a). To be specific, the testing effect refers to the phenomenon in which testing decreases the amount of information forgotten relative to studying (Figure 1). Figure 1 shows the results of a classic study on the testing effect (Roediger & Karpicke, 2006a). Participants in this experiment who were tested immediately after learning information (prose passages) had greater recall of the information 2 days and 1 week after the initial learning session. This experiment highlights the key findings that most of the testing effect literature finds, that while studying produces greater immediate retention, testing protects against forgetting and therefore leads to greater retention long term.
The benefits of testing are abundant, especially in education. For instance, testing aids in identifying gaps in a student’s knowledge. Testing a student’s knowledge allows one to know if the student adequately knows and understands the material. When given the opportunity, students often study missed items on tests more than the items they got correct, suggesting that students themselves use tests as a tool to guide their study habits (Son & Kornell, 2015). In fact, after an opportunity to study after an initial practice test, subjects corrected their mistakes on a second test (Amlund, Kardash, & Kulhavy, 2006).

Figure 1: Results from Experiment 1 (Roediger & Karpicke, 2006a).
Inopportune, students often do not self-administer practice tests as a form of study. Instead, students simply reread their material because rereading requires less effort and quickly leads to increased feelings of fluency (Karpicke, Butler, & Roediger, 2009). This feeling of fluency creates a false sense of knowledge and highlights the importance of utilizing practice tests over exclusively re-reading material.

Testing has also been shown to benefit students by increasing the organization of knowledge of the learned information. Organization of knowledge refers to how an individual stores and codes learned information. One example of how organization impacts learning is with recoding, or chunking (Miller, 1956). Chunking is the breaking down of large amounts of information into smaller parts that are easier to remember. For example, chunking often increases with phone numbers. The phone number 6166488113 may be chunked into three smaller parts as 616-648-8113. Increased organization of knowledge was thought to be one of the reasons first proposed for the benefits of testing (Gates, 1917) and testing has been shown to increase the organization of learned material (Zaromb & Roediger, 2010).

When studying, it can be difficult to measure the amount of information one possesses, compared to how much they think they know. Students often reread material increasing their familiarity of the material and their confidence of the knowledge of the material. This confidence is misleading and can cause students to incorrectly assess their amount of knowledge. Testing informs students of the knowledge that they actually possess (Roediger & Karpicke, 2006b; Shaughnessy, Zechmeister, & Zechmeister, 2013).

An additional benefit of testing is a reduction of proactive inference. Proactive interference can occur when previously learned material negatively affects the retention of new learned material (Crowder, 1976). When students are tested in between the learning of different
materials, testing prevents proactive interference from the previously learned material (Szpunar, McDermott, & Roediger, 2007). Testing has the potential to be a great tool for students to use to both allow students to assess their knowledge of the learned material and more effectively learn new material.

Teachers can use testing to overcome some challenges they may face in the classroom. One issue that teachers face is that they often overestimate the knowledge of their students (Kelly, 1999). Giving tests in the form of regular quizzing can allow teachers to assess the knowledge of the students and therefore teach accordingly (Black & Wiliam, 2010). Having to take regular quizzes also causes students to need to study more frequently (see Bangert-Drowns, Kulik, & Kulik, 1991; Leeming, 2002). Research suggests that students knowing that they will have a final test will outperform students who are unaware of a final test (Szpunar et al., 2007). Overall, testing has been shown to be an advantageous tool for teachers to use to increase the learning of their students.

The effect that testing has on learning has been studied extensively in real world classroom settings as well. For example, college students in a chemistry course were given multiple choice practice tests with the names of compounds and when tested a week later, those that completed the practice test performed better than those that did not (Yiğit, Kıyıcı, & Çetinkaya, 2014). In another example, students either took weekly quizzes followed by either short answer or multiple-choice questions or additional readings (McDaniel, Anderson, Derbish, & Morrisette, 2007). The students who answered additional short answer questions performed better on the final exam compared to those who answered additional multiple-choice questions and both groups performed better than those who only completed additional readings. The beneficial impact that testing has on student performance has been thoroughly studied in
psychology courses showing a significant increase in learning when tests are given (Schwieren, Barenberg, & Dutke, 2017). The positive effect that testing has on performance on exams has been replicated numerous times in the classroom and in the laboratory (e.g. McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; McDaniel et al., 2007; McDaniel, Wildman, & Anderson, 2012; Roediger, 2013; Schwieren et al., 2017).

Previous research on the testing effect has been limited in scope in that it has only focused on how testing impacts the learning and retention of declarative memory. This dissertation expands the testing effect literature by investigating the impact that testing has on materials thought to require nondeclarative memory. First, evidence for multiple memory systems are presented. Second, methodologies in which the testing effect have been studies are detailed. Third, multiple recent theories that attempt to explain the testing effect are discussed. Fourth, methodologies to study the testing effect with nondeclarative memory are proposed.

**Memory Systems**

Memory is not a single construct, instead, memory involves distinct systems that are functionally and neurologically dissociable. The two systems of focus for this proposal are the declarative memory system and the nondeclarative memory system (Squire, 1992). A key question presented in this dissertation is whether testing may impact memory differently depending on which memory system is recruited during learning.

**Declarative Memory**

Declarative memory can be thought of as the acquisition, retention, and retrieval of information that can be intentionally and consciously recollected, supported by the hippocampus and medial temporal lobe structures (Cohen & Squire, 1980). This information includes memories of specific events and facts (Tulving, 1983). Declarative memory is often measured
using direct tests of memory. Direct tests require consciously recollecting previously learned information (MacLeod & Daniels, 2000). These include free recall, cued recall, recognition, short answer, and multiple-choice (Graf & Schacter, 1985). Studies of the testing effect exclusively use tests of declarative memory.

Arguably the primary reason that declarative memory has been a focus is because the testing effect has often been explored in attempt to understand optimal learning strategies for students. Students often are tasked to memorize large amounts of information and recall it later during an examination or test in a classroom; therefore, it is important to investigate how testing impacts retention for this kind of memorization. The memorization of information is not the only type of learning that students engage in however. Students are often tasked to learn skills that do not only require declarative memory. Take learning how to complete long division for example. Students learn the rules of doing long division, and then practice often. Over time they become quicker to solve long division problems and become more accurate. This increase in accuracy and decrease in completion time reflects nondeclarative learning. Because of this, it is important to understand how testing impacts the learning and the long-term retention of these types of skills as well.

**Nondeclarative Memory**

Nondeclarative memory, also known as implicit memory, includes the acquisition, retention, and retrieval of knowledge that is expressed though changes in performance due to experience supported by different brain systems including the dorsal striatal brain systems for procedural memory, the neocortex for priming, and the cerebellum for different types of conditioning (Packard, Hirsh, & White, 1989; Squire & Dede, 2015). This type of memory is often measured by indirect tests of memory. An indirect test of memory is a test that does not
require the conscious recollection of previously learned information (MacLeod & Daniels, 2000). An example of this in an educational setting can be seen in chemistry classes (Taber, 2014). In these classes, students often must learn to use different instruments and tools which requires nondeclarative memory. The use of different tools is thought to require nondeclarative memory because something like pipetting a compound into a flask may not require conscious thought and may instead be nonconscious.

Nondeclarative memory is made up of procedural memory, the perceptual representation system (priming), classical conditioning, and nonassociative learning. Procedural memory depends heavily on extensive and repeated experiences and involves skills such as learning to ride a bike and more cognitive skills such as reading. Priming is the change in the response to something as a result of a prior experience with that stimulus and acts within the perceptual representation system (Tulving & Schacter, 1990). In classical conditioning, a conditioned stimulus is paired with an unconditioned stimulus (Pavlov, 1960). After the conditioned stimulus becomes associated with the unconditioned stimulus, the conditioned stimulus will evoke a conditioned response that is similar to the response typically evoked from the unconditioned stimulus. Nonassociative learning comprises habituation and sensitization (Groves & Thompson, 1970). Habituation is where a response to an unchanging stimulus decreases over time and sensitization is where a response to a stimulus increases over time with repeated presentation of that stimulus.

This dissertation focuses on the testing effect with specific components of nondeclarative memory. Experiment 1 and 2 use a word fragment completion task, which involves priming specifically (Roediger & McDermott, 1993; Roediger, Weldon, Stadler, & Riegler, 1992). Experiment 3 uses the weather prediction task which requires procedural learning to perform
optimally (Gluck, Shohamy, & Myers, 2002; Knowlton, Mangels, & Squire, 1996; Knowlton, Squire, & Gluck, 1994). In this dissertation, the term “declarative memory” is used to refer to the explicit, conscious, memorization of material. The term “nondeclarative memory” is used to refer to multiple aspects of nondeclarative memory, whether it is the priming task (word fragment completion) in Experiment 1 and 2 or the procedural learning task (the weather prediction task) in Experiment 3. This nomenclature is used for the sake of consistency. The differences between the types of nondeclarative memory in Experiments 1, 2, and 3 is explored in the discussion sections of the experiments in the dissertation.

**Evidence for a Dissociation**

In order for there to be a difference in how testing impacts declarative and nondeclarative memory, these two types of memory need to be fundamentally different in some way. Dissociations between declarative and nondeclarative memory have been shown both neurologically and functionally. In this next section, an argument for a dissociation is made. First, evidence is shown with nonhuman animal studies. Next, arguments are made for a similar dissociation in humans, using evidence from neuroimaging and neuropsychological studies.

**Dissociation – Animal Studies**

Perhaps the greatest evidence of a dissociation between declarative and nondeclarative memory systems comes from research on nonhuman animals. One of the, if not the best, ways to show a double dissociation is with a causal experiment that involves damaging multiple brain regions and showing differential effects on two different processes or tasks. Other than techniques like transcranial magnetic stimulation, this is impossible to do in humans. Invasive techniques such as causing lesions cannot be done in humans, and even in the case of patient studies, the subcortical areas damaged cannot be chosen. Even with transcranial magnetic
stimulation, subcortical regions cannot be targeted. Nonhuman animal studies allow for more causal evidence with more invasive techniques than with human studies.

Dissociations between a declarative and nondeclarative memory system have been shown in rats by examining how damaging brain regions impacts memory. For example, the Morris water maze (Morris, 1981), a task that can be manipulated to tap into both declarative and nondeclarative memory systems, has been used to investigate the ability of rats with either dorsal striatal or fornix damage to learn the location of a platform in a water maze (McDonald & White, 1994). Rats can learn to navigate this maze and swim to a raised platform by using stimuli in the environment as cues to the location of the platform. Rats with hippocampal damage (specifically the fornix, a part of the hippocampal formation and a brain structure often associated with declarative memory) have been shown to be impaired when the platform is submerged. With the platform submerged, and therefore not visible to the rats, the rats must use spatial cues present in the environment to locate the platform (Sutherland & Rudy, 1988). Rats with dorsal striatal damage, a brain structure often associated with types of nondeclarative memory, have troubles locating a platform when the platform is visibly moved. Instead, they swim to the location they had learned based upon the visual cues in the environment. Similar work with water mazes and lesions have replicated these results as well, further strengthening the argument for multiple memory systems (Packard & McGaugh, 1992; M. Packard et al., 1989). This shows a double dissociation, where the dorsal striatum is important for using environmental cues and where the hippocampus is important for learning the spatial location of the platform.

**Dissociation – Human Studies**

A case has been presented for a difference in both function and neurological regions for declarative and nondeclarative memory in nonhuman animals. Because the focus of this
dissertation is solely with humans, it is a necessity that these differences exist in humans as well.

A dissociation between declarative and nondeclarative memory will be argued for in the following section using evidence from neuroimaging studies, neuropsychological studies, and studies showing differences in function.

**Neuroimaging**

Research using neuroimaging has shown physical differences between the declarative and non-declarative memory system. For example, the declarative and nondeclarative memory systems have been shown to compete and interact in tasks that can recruit either system, for example the weather prediction task (Poldrack et al., 2001). Different memory systems are engaged depending on which version of the task is completed. During one version of the task, a feedback-based version, the basal ganglia is shown to be engaged (see Figure 2). Participants in the feedback-based version learn the association between geometric cues and weather patterns via trial and error. In another version of the task, a paired-associates version, participants learn this association by being shown the geometric cues at the same time as the weather patterns. In this version, the medial temporal lobe is engaged. Different neurological systems are differentially engaged depending on whether the task was learned incrementally via trial and error learning or if it was learned with more explicit memorization.
Evidence for a dissociation could be demonstrated if the effect of learning under dual-task conditions affects a declarative memory task and not a nondeclarative memory task. In one such study, the effect of learning under dual-task conditions was studied using fMRI (Foerde, Knowlton, & Poldrack, 2006). Participants learned two probabilistic category learning tasks under either single-task or dual-task (tone-counting) conditions. Nondeclarative memory is associated with performance that does not require effortful attention while declarative memory tasks do and therefore are more sensitive to a presence of a dual-task that also require effortful attention (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996). While percent correct did not differ between the two conditions, patterns of brain activity were different depending on the condition of the task. Accuracy in the single-task version correlated with activity in the medial
temporal lobe and accuracy in the dual-task version correlated with activity in the striatum. This is because declarative and nondeclarative learning mediates task performance. The declarative memory system, modulated by the medial temporal lobe, is more sensitive to concurrent distraction and this distraction can bias this competition so that the striatum takes over.

In line with the research showing that dual-task conditions can modulate the engagement of these memory systems, stress can also modulate these systems (Schwabe & Wolf, 2012). While stress induced prior to a probabilistic learning task does not affect the actual classification performance, it changes the nature of the learning. When stressed, participants remember less of the declarative, explicit aspects of the task itself and more often use a procedural, nondeclarative strategy. In stressed individuals, success in the probabilistic learning task was correlated with striatal activity. In nonstressed controls success was correlated with the hippocampus. Stress differentially affects the engagement of the memory systems and the learning strategy used.

Even something as small and simple as changing the instructions of a task can affect patterns of brain activity and which system is recruited. Using dot patterns (see Figure 3), a different pattern of brain activity is found depending on how the task was learned (Reber, Gitelman, Parrish, & Marsel Mesulam, 2003). When tasked to learn to categorize these dot patterns with explicit instructions (declarative learning), increased activity in areas like the hippocampus, right prefrontal cortex, and left inferior prefrontal cortex are shown. When simply shown the patterns, without mention of the category (nondeclarative learning), decreased occipital activity was observed. This suggests a dissociation of brain activity depending on whether a declarative or nondeclarative memory system is recruited. These results are especially important for this dissertation because in Experiment 1, the memory system recruited are manipulated by also changing the instructions of the task.
This research presented compelling evidence for a dissociation between declarative and nondeclarative memory systems. These memory systems can become differentially recruited showing differences in which regions of the brain are recruited by the task itself (Poldrack et al., 2001), by the inclusion of a secondary distractor task (Foerde, Knowlton, & Poldrack, 2006), and by differences in how the material is learned (Reber et al., 2003). This dissertation takes advantage of this ability to recruit different systems. In Experiment 1 and 2, using a word fragment completion task, participants learned a series of words either incidentally or intentionally, recruiting either nondeclarative or declarative memory processes. Experiment 3 used the same task as Poldrack and colleagues (2001) to recruit the declarative or nondeclarative memory systems.

Figure 3: Stimuli used in (Reber et al., 2003). (A) The prototype to be learned. (B) Low-distortion pattern. (C) High-distortion pattern. (D) Prototype of an unfamiliar category.
Neuropsychological

Some of the earliest research suggesting a dissociation between the declarative and nondeclarative memory systems comes from patients with neurological damage. One of earliest examples of a dissociation between these memory systems comes from the patient HM. As a treatment for epilepsy, HM underwent bilateral excision of substantial portions of his temporal lobes, affecting the hippocampus and related structures (Milner & Penfeild, 1955). HM’s memory was severely limited after the surgery and was unable to recall any experiences after they occurred (Milner & Penfeild, 1955). Though his declarative memory was impaired, he showed fewer impairments for his nondeclarative memory. He was able to learn to solve simple mazes and acquire a conditioned eye blink response, both of which require nondeclarative memory systems (Milner, Corkin, & Tuerber, 1968; Woodruff-Pack, 1993).

Patients with amnesia and individuals with striatum disorders (such as Parkinson’s or Huntington’s) are often targets for this type of memory research due to their ability to perform well in tasks that require one type of memory and not the other. Individuals with amnesia can perform relatively well in tasks that require nondeclarative memory whereas individuals with striatum disorders can do relatively well in tasks that require declarative memory. In one such study of memory, subjects were trained to read words that were presented in mirror-reverse text and their reading speed was measured, a measure of nondeclarative memory (Cohen & Squire, 1980). Healthy controls were able to increase the speed at which they were able to read the words and were able to correctly recognize which words they had seen during training. Subjects with amnesia due to Korsakoff’s syndrome were able to increase their reading speed but were impaired at recognizing the words they had seen. An opposite pattern has been observed in subjects with a damaged striatum due to Huntington’s disease (Martone, Butters, Payne, Becker,
& Sax, 1984). These subjects had no issue remembering the words presented (a test of declarative memory) but were impaired at learning the mirror-reading skill. These experiments show a double dissociation of declarative and nondeclarative memory.

Individuals with amnesia who have shown impairment on a declarative memory task perform normally on nondeclarative memory tasks like the word fragment completion task (Blaxton, 1992; Cermak, Blackford, O’Connor, & Bleich, 1988; Graf, Squire, & Mandler, 1984; Hamann & Squire, 1997). The word fragment completion task is used as a test of nondeclarative (priming) memory (Graf & Schacter, 1985; Roediger & McDermott, 1993; Roediger et al., 1992; Larry R Squire, 1987). Priming is the change in how an individual identifies or produces an item as a result of prior experiences with that item (Tulving & Schacter, 1990). Generally, in word fragment completion tasks, participants study words and then are shown fragments of both studied and unstudied words (Roediger et al., 1992). For example, a participant is instructed to complete the word fragment “c_m_u_t_r” (for “computer”) with the first word that comes to mind. The word completion task is considered a perceptual priming task and requires that stimuli are processed based on their physical characteristics, not their semantic meanings (Blum & Yonelinas, 2001; Kinjo & Snodgrass, 2000; Roediger & McDermott, 1993; Rueckl & Mathew, 1999). In fact, neuroimaging studies have shown that the word fragment completion task is associated with areas linked with nondeclarative memory, such as the extrastriate cortex, whereas declarative memory is more associated with the medial temporal lobe (Schacter, Alpert, Savage, Rauch, & Albert, 1996; Schott et al., 2006).

The first and second experiments utilize a word fragment completion task to investigate how testing impacts long term retention for tasks that require nondeclarative memory. The effect that being tested on these word fragments has on the long-term priming is currently unknown.
Experiments 1 and 2 of this dissertation investigate this using the word fragment completion task (a nondeclarative memory task) and a declarative memorization task where participants are asked to explicitly memorize the words. Based on previous research with the testing effect showing a strong testing effect when participants are asked to memorize words (Carpenter & Delosh, 2006; Hogan & Kintsch, 1971; McDaniel & Masson, 1985; Rowland, 2014; Endel Tulving, 1967; Zaromb & Roediger, 2010), a testing effect is expected with the declarative memory condition of this task. It is currently unclear if a testing effect will be found with the nondeclarative memory condition of the word fragment completion task.

One other example where patients with amnesia and individuals with striatum disorders differ on performance is with the weather prediction task (Knowlton, Mangels, Squire, 1996; Knowlton, Squire, & Gluck, 1994). The weather prediction task was originally designed as a probabilistic learning task (Gluck and Bower, 1988). The probabilistic nature of the task is thought to make it difficult to memorize the associations between the geometric cues and the weather patterns. Instead, this type of task may cause a bias towards a nondeclarative memory system that learns via trial and error (Knowlton et al., 1994). During the task, subjects were required to learn which of two weather outcomes, sun or rain, was predicted by different combinations of one, two, three, or four different cues. The cues were four different cards, each with a different geometric cue. There were 14 total patterns made up of different combinations of these cards. Each cue was associated to either weather outcome with a fixed probability. Subjects were instructed that they would be shown one, two, three, or four cues with geometric symbols on each trial and they should decide if the cues presented predicted sun or rain. During each trial, subjects were presented with feedback in the form of a smiling face if they were correct, and a frowning face if they were incorrect as well as the correct weather pattern.
Patients with amnesia have been shown to learn the weather prediction task at similar rates to healthy controls for the first 50 trials, with an impairment relative to controls in later trials (Knowlton et al., 1996, 1994). The reason that the task is learned at similar rates during the first 50 trials, is that the first 50 trials are almost solely dependent on nondeclarative memory processes. Later trials may benefit from declarative memory processes, which would explain why controls perform better on later trials.

Patients with Parkinson’s disease or Huntington’s disease, who have basal ganglia dysfunction, show impaired learning of the relationship between the weather patterns and the cues from the start of training, but performed normally on a test that assessed declarative memory of the task (Knowlton et al., 1996). Patients with schizophrenia demonstrate similar categorization performance as controls, with impairments on the recognition of the category cues (Kéri, Antal, Szekeres, Benedek, & Janka, 2000). This is consistent with previous studies suggesting individuals with schizophrenia show impaired declarative, not nondeclarative memory. Overall, the weather prediction task has been used to show a clear dissociation between the memory systems.

In Experiment 3, the weather prediction task is used to investigate the difference in the magnitude of the testing effect for the different memory systems. The paired-associates version of the weather prediction task should encourage declarative memory strategies whereas the feedback version should encourage nondeclarative memory strategies (Poldrack et al., 2001). Participants learned to complete the weather prediction task with either the paired-associates or the feedback version of the task. Participants completed a final evaluation of their memory either immediately after training or 48 hrs later. The magnitude of the testing effect is measured by performance on the final evaluative test. A greater testing effect would be shown by a smaller
decrease in performance from the first test to the second test, 48 hours later, in the feedback version (testing) as compared to the paired-associates version (study).

Conclusion

Using examples from nonhuman animal, neuropsychological, and neuroimaging studies, a dissociation between declarative and non-declarative memory systems is presented. The different memory systems are supported by different physical structures, the declarative supported by the hippocampus and medial temporal lobe systems and the nondeclarative supported by dorsal striatal brain systems for procedural memory, the neocortex for priming, and the cerebellum for different types of conditioning (Packard et al., 1989; Squire & Dede, 2015). The testing effect has only been studied thus far with declarative memory tasks. This dissertation is the first to investigate the effect that testing has on retention for nondeclarative memory tasks using the word fragment completion task and the weather prediction task.

This dissociation between declarative and nondeclarative memory systems is important for a number of reasons. One reason is different memory systems benefit more from different methods of learning. For example, an elementary student tasked to memorize a map of the United States would recruit a declarative memory system whereas a student learning to use some computer software may recruit a procedural, nondeclarative memory system. Different study techniques may be more beneficial depending on which type of task is being learned. Previous research has shown that testing is an effective way to learn tasks that require declarative memory (Roediger et al., 2006a), but it is currently unclear if testing shows the same benefit for tasks that require nondeclarative memory. This dissertation is the first to investigate the effect that testing has on nondeclarative memory tasks.
While research shows that the testing effect is a robust phenomenon, it should be noted that there are often substantial amounts of variance in the strength of the testing effect. The strength of the testing effect depends on numerous factors such as where the information is learned, the type of information learned, the time delay between learning it and being tested, and the type and amount of feedback. Differences in methodologies are discussed below followed by a discussion on the methodologies that produce the greatest testing effect. Since this dissertation is the first to use nondeclarative memory tasks to investigate the testing effect, one aim was to use a methodology that could promote a testing effect in nondeclarative memory by adapting methodologies from tasks that have shown the testing effect in declarative memory tasks.

Testing Effect Methodologies

Within the laboratory, the testing effect is generally studied using an initial study phase, a training phase, and a final evaluation phase (see Roediger & Karpicke, 2006b for a review). The initial study phase involves having participants learn some sort of material, often word-pair associations or information contained in prose passages. The training phase, also known as the practice phase or the re-exposure phase, typically involves either presenting the learned material an additional time (known as the study condition) or testing participants on the learned material (known as the testing condition). Lastly, the final evaluation phase involves a final test of the initial learned information and acts as a measure of retention. While the positive effect that testing has on retention has been shown to be a robust phenomenon, the magnitude of the testing effect depends on aspects of the methods used such as the testing environment, the experimental design of the study, whether or not there is feedback, the type of feedback, and the comparison treatment used. (Carpenter & Pashler, 2007; Marsh et al., 2007; Roediger & Karpicke, 2006b). Because using nondeclarative tasks to investigate the testing effect is novel, it is important to
employ an experimental design which would lead to the highest probability of detecting a testing effect. This is assuming that testing operates in a similar way for declarative and nondeclarative memory. If a testing effect is not found, it is imperative that it is not a result of the methodology. Below various methodologies used to investigate the testing effect are detailed and the best methodology for this dissertation is argued for.

**Comparison Conditions**

When investigating the testing effect, there is usually a test condition and a comparison condition. One area where testing effect studies differ from each other is with what kind of comparison condition is used. The test condition involves participants completing a practice test (during the training phase) after the initial learning phase and before the final evaluation phase. The performance on the final evaluation is compared between this testing condition, and a comparison (study) condition. The most common types of comparison treatments are re-studying, re-reading, and the use of a filler-task or no activity.

Most testing effect experiments use a comparison condition that involves completing the initial study phase an additional time (during the training phase). One reason for this is to reconcile the otherwise potential confound that testing increases long term retention only because those tested have an additional chance to process the material. By presenting participants in the study condition the material an additional time, participants in both the test condition and the comparison (study) condition are exposed to the material the same amount eliminating that confound. Also, having participants in the comparison (study) condition simply re-study the initially learned material mimics what students often do to prepare for tests, read over their notes. While most testing effect studies use a comparison condition where participants are exposed to
the materials again, in some studies the participants in the comparison condition complete no additional activity (e.g. Wheeler & Roediger, 1992).

In a re-study condition, participants often simply complete the initial study phase (during the training phase) an additional time (e.g. Abel & Roediger, 2018; Tulving, 1967). This additional study phase is usually identical to the initial learning phase. For example, Abel & Roediger (2018) presented vocabulary word-pairs to participants one at a time during an initial learning phase. Participants in the comparison (study) condition had participants re-study the word pairs one at a time identical to the initial learning phase. A testing effect was found in that there was less forgetting for the words that were tested during a practice test as compared to words that were studied again. A re-reading comparison condition acts functionally similar to a re-study comparison condition except that participants in a re-reading comparison condition learn materials by reading passages during both the initial study phase and the training phase (e.g. Kornell & Bjork, 2007; Kornell & Son, 2009; Roediger & Karpicke, 2006a). For example, Roediger & Karpicke (2006a) had participants study a passage for 7 minutes during the initial learning phase and those in the rereading comparison condition re-read the same passage during the training phase. Re-studying and re-reading are the most common forms of a comparison condition (Adesope, Trevisan, & Sundararajan, 2017).

While re-reading and re-studying are the most common forms of a comparison condition, it should be noted that using a filler task has been shown to have on average a greater effect size (Hedges’ g = 0.93) as compared to re-studying (0.51), (Adesope et al., 2017). As mentioned, a comparison condition is used to equate the exposure of the to-be-remembered information. Without a comparison condition, it is difficult to argue that testing is an effective learning

---

1 Hedges’ g is a measure of effect size that adjusts the effect size to create an unbiased estimate of the effect size (Freeman, Hedges, & Olkin, 1986; Hedges, 1981).
strategy and not just an additional practice opportunity. Therefore, the experiments in this dissertation used re-study for the comparison conditions in order to equate time and content between a testing and a study condition.

**Testing Environment**

Another important design choice when investigating the testing effect is the testing environment. The testing effect has been studied in both laboratory settings and classroom settings and each has their own specific benefits. Laboratory research is the more common approach (accounting for 89% of testing effect studies; Adesope, Trevisan, & Sundararajan, 2017), where participants come in, are trained, and then tested in the same environment (e.g. Roediger & Karpicke, 2006a). The obvious benefit of this approach is that you can test participants in a very controlled environment. The researcher can choose what the participants are trained on, how they are trained, how much time exists between the training phase and the testing phase, and the populations to be studied. This potentially presents a problem for the ecological validity however, since in the real-world students are not asked to learn in such a controlled environment.

The testing effect has been shown in both classroom and laboratory settings. In fact, the effect size for classroom studies (g = .67) and for laboratory settings (g = .62) are similar. For the purpose of this dissertation, a laboratory setting was used. The reason is to control things like the length of the experiment, the surrounding environment, and the instructions given to the participants.

**Experimental Design**

Studies of the testing effect employ either a within-subjects design, a between-subjects design, or a mixed-design. Between-subjects designs often use a training type by timing research
design where participants are split based on both the type of training and the delay period between training and test (e.g. Karpicke & Blunt, 2011; McDaniel & Fisher, 1991). One drawback for a between-subjects design is that it requires twice as many participants as a within-subjects to obtain the same power to detect an effect. Also, unless controlled for, prior knowledge between the groups can add as a source of error. Random assignment can also help eliminate between group differences. In a within-subjects design, some of the stimuli learned during the initial training are re-studied and the other stimuli are tested during the training phase (E.g. Kornell, Bjork, & Garcia, 2011; Pyke, Bourque, & LeFevre, 2019; Storm, Friedman, Murayama, & Bjork, 2014). This design eliminates the potential confound of prior group differences that could be present in a between-subjects design. One potential negative however is that the testing of some items could impact the retention of other non-tested items (Chan, 2009, 2010). Mixed-designs incorporate both the different training and timing conditions as well as having some of the trained items tested and some re-studied (e.g. Abel & Roediger, 2018; Carpenter et al., 2006; Roediger & Karpicke, 2006a).

Since the testing effect has been shown in between-subjects, within-subjects, and in mixed designs, the experiments in this dissertation used whichever is appropriate given the experimental paradigm. In Experiment 1 and 2, using a word fragment completion task, a mixed-subjects design was used. The participants were trained by re-studying some of the words and testing the others (within-subjects). Keeping this within-subjects eliminates prior group differences so any difference in memory between tested and studied items will be a result of training. The participants were randomly assigned into one of two timing conditions, immediate or delay (between-subjects). Experiment 3, using the weather prediction task, used a between-subjects design only. The reason for this is that testing on some of the patterns could affect the
memorization of other non-tested items. Like Experiment 1, the timing conditions were between-subjects.

**Feedback**

Feedback refers to the section of the experiment in which the participant’s answers to a practice test are indicated as correct or incorrect. While feedback is not necessary for the testing effect to be observed, the presence of feedback as well as the type of feedback affects the strength of the testing effect. Presenting feedback in classroom settings have been shown to enhance performance on exams (Butler & Roediger, 2008; McDaniel & Fisher, 1991). Feedback that is presented immediately after a response has been shown to lead to higher final exam scores in classroom settings (Bangert-Drowns et al., 1991). In some cases, however, feedback has been shown to be ineffective. Feedback presented after high-confidence answers may be an inefficient use of time and cognitive energy (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Hays, Kornell, & Bjork, 2010). If a person is highly confident that they are correct, then feedback gives no additional information. The impact that feedback has on retention and the magnitude of the testing effect depends on the difficulty of the material learned, how well it is learned, the amount of time after a response, and the confidence of the given answer.

Feedback is especially important when considering nondeclarative, procedural learning. In fact, feedback is necessary for procedural learning (Schultz, 1992; Wickens, 1993). Feedback acts as an unexpected reward that causes a dopamine release, which then strengthens recently active synapses. This importance of feedback can be seen in the weather prediction task. In fact, Poldrack and colleagues (2001) found that by simply changing the task from a feedback version to a paired-associates version (a version where the correct version is displayed concurrent with the cards), neural systems implicated in declarative learning became more active whereas
systems implicated in procedural learning became less active. Therefore, feedback was used in the feedback version of the weather prediction task in Experiment 3.

Other types of nondeclarative learning are less feedback dependent. Tasks like the word fragment completion task often have participants rate a list of words on their valance (for example see Roediger et al., 1992). Participants are not told to memorize the words and therefore it is expected that they do not consciously try and memorize the words. When presented with a fragment of the word during a test phase, they are more likely to complete the fragment than if they had not rated the word earlier. This is an example of priming, a type of nondeclarative learning. In fact, there is reason to believe that feedback could be detrimental to the word fragment completion task and cause a declarative memory system to be recruited instead of a nondeclarative memory system (Graf et al., 1984). Therefore, feedback was not used in the word fragment completion task in Experiment 2.

**Delay Between Training and Testing**

Since information is often learned to be used at a later date, the delay between an initial learning phase and a test phase is important. Consider a student in a classroom. They are learning information to be used on some cumulative final exam that could occur a week later, a month later, or three months later at the end of a semester. In order for them to choose the most beneficial study technique, they would want to study in such a way that would have the greatest benefit on long term retention. The general idea behind the testing effect is that testing enhances retention of items by decreasing forgetting. This decrease in forgetting may not be apparent if there has not been adequate time for the items that were learned to be forgotten. Many researchers have found that there is either no testing effect or even a reverse testing effect at short retention intervals (Bouwmeester & Verkoeijen, 2011; Congleton & Rajaram, 2012;
However, some studies have shown an enhancement of testing relative to study in the same order of minutes (e.g., Carpenter & Delosh, 2006; Carpenter & DeLosh, 2005; Karpicke & Zaromb, 2010; Rowland & DeLosh, 2014). In many cases, a testing effect by retention interval interaction has been shown. As the retention interval increases, the testing effect also increases (e.g., Congleton & Rajaram, 2012; Verkoeijen, Bouwmeester, & Camp, 2012). This delay/testing effect interaction is important to consider when designing a testing effect study. While longer delays increase the strength of the testing effect, a delay of only 2 days is more than enough to show a significant testing effect (Roediger et al., 2006a). Therefore, a delay of 2 days was used for the experiments in this dissertation.

**Format of Practice Test**

The actual format of the practice test during the training phase varies between experiments of the testing effect. Some utilize free-recall tests by presenting some information during the initial study phase and then having participants recall all of the information possible in a given time (e.g. Hogan & Kintsch, 1971; Lachman & Laughery, 1968; Roediger & Karpicke, 2006a; Tulving, 1967; Wheeler & Roediger, 1992). Others use cued-recall and require participants to generate information by giving participants a cue to aid in recall (e.g. Carpenter, Pashler, & Vul, 2006). One type of cued-recall, short answer, prompts the participants with a question and the participant must produce the answer (e.g. Chan, McDermott, & Roediger, 2006).

Despite differences in the actual format of the practice test, strong testing effects are still shown. A recent meta-analysis of the testing effect literature calculated differences in Hedges g for the strength of the testing effect for different formats of the practice test (Adesope et al.,
Mixed-format tests had the strongest effect (g = .80; though the total number of studies with this type of format was only 2). The next highest formats were multiple-choice (g = .70), free-recall (g = .62), cued-recall (.58), and short-answer (g = .48). Each format of practice tests produces at least a medium effect size. Experiment 1 and 2 used a type of cued recall (word fragments) and Experiment 3 used cued recall where participants are cued with a geometric shape and must respond with the correct response.

Test Materials

Like the format of the practice tests, the actual test materials vary between testing effect experiments. The most common materials used are single word lists, paired-associates, prose passages, and nonverbal materials. When single word lists are used, lists of words are presented, one at a time, for between 3-7 seconds a word (Carpenter & Delosh, 2006; Hogan & Kintsch, 1971; McDaniel & Masson, 1985; Rowland, 2014; Endel Tulving, 1967; Zaromb & Roediger, 2010). Participants are instructed to remember these words for a later memory test. Words that are tested during the training phase are recalled at greater amounts than those studied. Paired associate tests are similar to single word lists except participants are asked to learn word-pairs instead of single words (Allen, Mahler, & Estes, 1969; Carpenter, 2009; Carpenter et al., 2006; Carrier & Pashler, 1992; Kornell et al., 2011; Pyc & Rawson, 2010; Toppino & Cohen, 2009). Experiment 1 and 2 used single word lists to investigate the testing effect.

Other experiments utilize prose passages to present information to participants (e.g. Chan et al., 2006; Glover, 1989; Roediger & Karpicke, 2006a; Spitzer, 1939). For example, Roediger and Karpicke (2006a) used two prose passages taken from the reading comprehension sections of a test-preparation book for the Test of English as a Foreign Language. Participants were instructed to read the two prose passages during the study phase and write down as much as they
could remember during the test phase. Roediger and Karpicke (2006a) found the classic testing effect. Information that was retrieved during a practice test was forgotten less as compared to the studied information.

Experiment 3, using the weather prediction task, used nonverbal materials. The weather prediction task involves participants learning the association between geometric cues and weather patterns. While there is some research showing that testing with nonverbal materials increases long term retention (for example see Carpenter & Pashler, 2007; Kang, 2010; Wheeler & Roediger, 1992), it is not as common as using verbal materials. Despite it being less common, research shows that using nonverbal materials is acceptable for measuring the testing effect.

**Summary**

The purpose of detailing the various methodologies was two-fold. One, it illustrates that despite their being such a wide variety of methodologies used, the testing effect remains a robust finding. The second, is to use the optimal conditions to produce the greatest testing effect for use in the experiments in this dissertation. Each experiment in this dissertation used a re-study comparison condition in order to equate the exposure of the tested and studied materials. The experiments took place in a laboratory, instead of a real-world setting like a classroom, in order to have the most control over aspects like the timing of the experiment, and the instructions given. All three experiments had an immediate condition and a delay condition that takes place two days later, since it has been shown that two days is enough time to show a testing effect.

Experiments 1, 2 and 3 each differ in both the feedback and the experimental design. Experiment 1, the word fragment completion task, utilized a mixed-subjects design in order to reduce error caused by prior group differences. Experiment 2 also did not give feedback during the training phase because doing so could cause the declarative memory system to be recruited.
instead of the nondeclarative memory system. Experiment 3, using the weather prediction task, used a between-subjects design due to the inability to split up the materials into test and study materials within participants. There was also be feedback in the nondeclarative condition due to the necessity of feedback for nondeclarative (procedural) learning.

Despite choosing methods that have been shown to produce testing effects in past research, it is currently unclear if testing enhances long term memory by decreasing forgetting for tasks that primarily recruit nondeclarative memory. If testing effects nondeclarative memory in the same way that it effects declarative memory, then a testing effect can be expected. The aforementioned methods were chosen to produce the greatest likelihood that a testing effect would be found, assuming that testing increases long term memory for nondeclarative memory.

**Testing Effect Theories**

A long history of research shows that testing enhances long term declarative memory retention by decreasing forgetting (Roediger & Karpicke, 2006a). The question is though, why? What about testing causes decreased forgetting? Is there one major mechanism that causes decreased forgetting after resting, or are multiple mechanisms at work? Would the mechanisms cause testing to increase retention with declarative memory tasks also cause decreased forgetting with nondeclarative memory tasks? Several different theories have been proposed that specify a mechanism that attempts to explain some aspect of the testing effect. These theories are not necessarily exclusive in that one theory can explain a portion of the testing effect that another cannot. Therefore, many of these should not be seen as competing theories.

In this next section, a number of the more prominent of these theories are described in detail. These include transfer appropriate processing (Morris, Bransford, & Franks, 1977), desirable difficulties (Landauer & Bjork, 1978) the distribution-based bifurcation model
(Halamish & Bjork, 2011; Kornell et al., 2011), and the dual systems theory (Rickard & Pan, 2018). First, these theories of the testing effect that could make predictions about the testing effect with nondeclarative memory are detailed. Second, for each theory, an argument is made as to whether the theory predicts a testing effect with nondeclarative memory. Finally, a general argument is made using these theories for why a testing effect would be shown in the tasks used in this dissertation.

**Transfer Appropriate Processing**

Transfer appropriate processing refers to the state-dependent learning between how information is encoded and how it retrieved (Lockhart & Craik, 1972). This theory has been used as a potential explanation for the testing effect (Morris et al., 1977). Typically, practice tests resemble the final test more than studying would. Imagine a student in a college course, throughout the semester they may take multiple tests in an attempt to be prepared for a final cumulative exam. One reason that these tests could help on the cumulative exam is they provide a way for the student to practice with questions similar in design to the cumulative final. If transfer appropriate processing is the primary explanation for the testing effect, then it would be expected that regardless of the type of test used, a match between a practice test and a final test would result in the greatest retention. This is not always the case. Practice tests with free recall type questions produce a greater testing effect than multiple choice and cued recall, regardless of the format of the final exam (Carpenter & Delosh, 2006). The reason for this may be that free recall tests require deeper memory processes than multiple choice testing (Kang, McDermott, & Roediger, 2007). Multiple choice tests may rely more on familiarity and therefore may have less positive transfer to a final multiple-choice test. So, while transfer appropriate processing may account for some of the testing effect, it does not provide complete account of the testing effect.
Since the main goal of this dissertation is to investigate the testing effect with nondeclarative memory tasks, an appropriate question is if there would be a similar benefit to retrieval with nondeclarative memory tasks. With nondeclarative memory tasks there is still a match between a practice test and a final test that does not exist between study and the final test. Therefore, according to the idea of transfer appropriate processing, there should be an enhancement of testing with nondeclarative memory due to this match that is only present when there is a practice test. This is especially important for Experiment 1 and 2 which use the word fragment completion task, a task that requires perceptual priming. Perceptual priming requires some physical match on primed information and the retrieved items, and a greater match between the words in the practice test and final evaluative test could lead to greater retrieval for the tested items (Blum & Yonelinas, 2001; Kinjo & Snodgrass, 2000; Roediger & McDermott, 1993; Rueckl & Mathew, 1999). It should be noted, however, that a match between a practice and final test is not required for the testing effect to exist, so it is possible that this match wouldn’t lead to a testing effect for perceptual priming (Carpenter & Delosh, 2006).

**Desirable Difficulties**

One reason that students may choose to read over their notes instead of engaging in practice testing is that testing requires greater perceived effort than studying. This relative difference in difficulty may be advantageous. A potential explanation for why testing increases retention compared to studying may actually be because the testing produces what is known as a desirable difficulty (Landauer & Bjork, 1978). By increasing the delays between the initial learning and testing, the difficulty of the practice tests increased. As a result, the amount recalled was increased during test phase. More difficult tests may cause individuals to have to work harder and engage in deeper processing, therefore increasing the amount of material retained.
Testing may provide a desirable difficulty that enhances the memory strength of the items in the testing condition (Landauer & Bjork, 1978). The first question that should be answered regarding a desirable difficulty is whether testing is more difficult than study. In the testing effect literature, practice testing often produces impaired retention when the final test is immediately after the practice test, suggesting that with the declarative memory tasks seen in the testing effect literature, testing is more difficult than study (Roediger & Karpicke, 2006a). Testing in a nondeclarative memory task, such as the word fragment completion task, can result in only 60% of fragments completed correctly for studied words (Rossi-Arnaud, Cestari, Rezende Silva Marques, Bechi Gabrielli, & Spataro, 2017). This proportion correct is much lower than seen in word-pair associate tasks seen in the testing effect literature (Roediger & Karpicke, 2006a) and, while difficulty is a not a direct result of proportion correct, it could be argued that testing with word fragment completion tasks maintains the difficulty present in the testing effect literature. Therefore, based on the desirable difficulties model, a testing effect is expected in nondeclarative memory tasks, including the word fragment completion task.

**Distribution-based Bifurcation Model**

The distribution-based Bifurcation Model attempts to explain why testing benefits retention relative to studying with a larger focus on the test-delay interaction relative to other theories of the testing effect (Halamish & Bjork, 2011; Kornell et al., 2011). The test-delay interaction refers to the idea that the testing effect has a greater effect with greater delays (Roediger & Karpicke, 2006a). Overall, this theory states that testing helps items recalled during testing to a greater degree than studying helps items that are only restudied. Restudying helps all items, just to a lesser degree than successfully retrieved items during test. Items that could not be successfully retrieved during a practice test are not aided by testing at all.
Figure 4: The distribution-based bifurcation model (Halamish & Bjork, 2011; Kornell et al., 2011). Each line represents memory strength at different time points for both restudy (A) and testing (B) conditions. Shaded areas represent the items recalled on the final test.

The way that this works is that testing creates a non-normal distribution of memory strengths (See figure 4). In figure 4, the top graph (A) represents a restudy condition and the bottom graph (B) represents a testing condition. Each distribution represents the memory strength for learned items at different time points. The left-most curve represents the memory strength for items before they are learned during an initial study phase. The middle curve represents the memory strength for items learned after an initial study phase and the right-most,
bold curve represents the memory strength for items after either a restudy period (A) or a testing period (B). The shaded areas represent the number of items successfully recalled during a final evaluation phase.

Tests bifurcate the distribution of item strengths. Items recalled successfully have greater strength, and items with greater strength are more likely to be recalled. Items are only recalled if its strength is above a recall threshold. This model makes specific predictions about the effects of feedback. Presenting feedback helps strengthen items that were not successfully recalled during testing, thus potentially eliminating, or at least suppressing, the difference in strength between successfully recalled and not successfully recalled items. According to the bifurcation model, the test delay interaction exists because of the bifurcation that testing without feedback creates. Thus, feedback eliminates the test delay interaction and the apparent prevention of forgetting that testing creates.

The distribution-based bifurcation model does not make predictions that are necessarily specific to declarative memory. If it is assumed that testing increases the item strength for tested items for nondeclarative memory tasks as well, then according to this model there would be a testing effect for nondeclarative memory tasks. And therefore, a testing effect would be expected for the word fragment completion task and the weather prediction task used in this dissertation.

**Dual Systems Theory**

The Dual Systems Theory argues for separate memories that are formed as a result from studying and testing (Rickard & Pan, 2018). The initial study phase creates a study memory and restudying strengthens this specific study memory. Initial testing, often in the form of a practice test, strengthens the study memory and encodes a new, completely separate test memory. Both the study memory and the test memory can support final test performance. Final testing in a
restudy condition can only be supported by the study memory whereas a final test after a practice
test can be supported by both the study and test memory. This difference is the cause of the
testing effect.

Test memory is made up of two components, a cue memory (the episodic encoding of the
retrieval cue in the present context) and an association between the cue memory and the correct
response. An association between the cue and the correct answer occurs when the correct answer
is retrieved from episodic study memory, or when feedback is provided on an incorrect trial.
Feedback on correct trials has no effect on final test performance (see Butler & Roediger, 2008;
Pashler, Cepeda, Wixted, & Rohrer, 2005). These components can form an additional route to
retrieval that only studying cannot form. Correct retrieval during test in the restudy condition is
thought to occur when the study memory strength is above some response threshold. Correct
retrieval during a final test in the testing condition occurs when the threshold is met for the study
memory, testing memory, or both. The Dual Systems theory is unique in that it is supported by a
model that predicts both proportions correct and the magnitude of the test effect (magnitude
referring to the greater difference in proportion correct as the time between the restudy/test and
the final evaluation phase increases).

The Dual Systems Theory states that separate memories are formed as a result from
studying and testing (Rickard & Pan, 2018). Final testing can be supported by both the study
memory created during the initials study phase and the testing memory created during the
practice test. One question to address would be if these two types of memory are specific to
declarative memory tasks. The word fragment completion task is thought of as a nondeclarative
memory task that requires priming (Tulving & Schacter, 1990). Like word-pair associate tasks
seen in a more typical testing effect study, a study memory could form during the study phase of
a word fragment completion task. If a testing memory is also created during a practice test for Experiment 1, using the word fragment completion task, then the Dual Systems Theory would predict a testing effect. Similar logic could be used in support for a testing effect with Experiment 3, using the weather prediction task.

**Other Theories**

The theories presented above can be used to make claims as to why there would be a testing effect in the experiments in this dissertation, but there are prominent theories worth mentioning that arguably make no claim for a testing effect with tasks that recruit nondeclarative memory. These are mentioned briefly as an overview of the testing effect literature.

The Elaborative Retrieval Hypothesis states that during testing, semantically related items are recalled, thus aiding in retrieval of some target item (Carpenter, 2009). The Mediator Effectiveness Hypothesis states that testing enhances long term retention by supporting the use of more effective mediators during encoding (Pyc & Rawson, 2010). The Gist Trace Processing theory focuses on individual differences in the testing effect and whether they can be explained by processing of fuzz representations of past events (Bouwmeester & Verkoeijen, 2011). The Episodic Context Theory proposed that retrieval enhances retention because people are required to think back to and reinstate a prior learning context (Karpicke, Lehman, & Aue, 2014). Finally, the Attenuated Error Correct Theory states that testing may create a more reliable error signal than studying, and this error signal is dependent on feedback (Mozer, Howe, & Pashler, 2004). Since feedback is not present in the conditions of the experiments that are thought to tap into nondeclarative memory, this theory would not make a prediction on the testing effect for them.
Summary

Each of these theories proposed make some claim about mechanism that may be important in understanding the testing effect. Some theories are expressed purely in conceptual terms without quantitative implementation for the testing effect, with Mozer et al.’s (2004) attenuated error correction theory and Rickard & Pan’s (2018) dual systems theory being the exception. These theories are not exclusive in that it is possible that the mechanisms described by the theories may contribute to the testing effect differently depending on the specific testing effect environment. While the testing effect is one of the most studied phenomena in cognitive psychology, there is still no consensus as to the exact mechanism to explain it.

An attempt has been made to use the prominent theories of the testing effect to hypothesize why a testing effect would be shown in tasks that recruit nondeclarative memory such as the word fragment completion task and the weather prediction task. The theory of transfer appropriate processing predicts a testing effect if the practice test is similar to the final evaluative test. Assuming that testing produces a desirable difficulty for tests that require nondeclarative memory (specifically procedural memory) similarly to declarative memory, the theory of desirable difficulties would predict a testing effect (Landauer & Bjork, 1978). If testing bifurcates memory strengths for tested and studied items for nondeclarative memory, the Distribution-based Bifurcation model would predict a testing effect as well (Halamish & Bjork, 2011; Kornell et al., 2011). Finally, if for nondeclarative memory, a separate test memory and study memory are created in a way similar to declarative memory, then the Dual Systems Theory would predict a testing effect for nondeclarative memory (Rickard & Pan, 2018). Overall, if nondeclarative memory operates in a similar way in regard to these theories, a testing effect could be expected for nondeclarative memory.
Table 1. Brief descriptions of different theories that attempt to explain the testing effect.

<table>
<thead>
<tr>
<th>THEORY</th>
<th>DESCRIPTION</th>
<th>CITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRANSFER APPROPRIATE PROCESSING</td>
<td>Refers to the state-dependent learning between a practice test and a test.</td>
<td>(Morris et al., 1977)</td>
</tr>
<tr>
<td>DESIRABLE DIFFICULTIES</td>
<td>Practice tests are more difficult than studying, causing a deeper processing of the information.</td>
<td>(Landauer &amp; Bjork, 1978)</td>
</tr>
<tr>
<td>DISTRIBUTION-BASED BIFURCATION MODEL</td>
<td>Testing increases the memory strength of tested items more than studying increases the strength of studied items.</td>
<td>(Halamish &amp; Bjork, 2011; Kornell et al., 2011)</td>
</tr>
<tr>
<td>DUAL SYSTEMS THEORY</td>
<td>Studying and testing creates two types of memory, study and test memory. Both types of memory can support a final test.</td>
<td>(Rickard &amp; Pan, 2018)</td>
</tr>
<tr>
<td>ELABORATIVE RETRIEVAL HYPOTHESIS</td>
<td>Items semantically related to the target are recalled during testing and aid in retrieval.</td>
<td>(Carpenter, 2009)</td>
</tr>
<tr>
<td>MEDIATOR EFFECTIVENESS HYPOTHESIS</td>
<td>Testing increases the amount of connections between the learned material.</td>
<td>(Pyc &amp; Rawson, 2010)</td>
</tr>
<tr>
<td>GIST TRACE PROCESSING THEORY</td>
<td>Individual differences in the testing effect can be explained by individual differences in processing of fuzzy representations</td>
<td>(Bouwmeester &amp; Verkoeijen, 2011)</td>
</tr>
<tr>
<td>EPISODIC CONTEXT THEORY</td>
<td>Prior learning contexts are reinstated during a final test.</td>
<td>(Karpicke et al., 2014)</td>
</tr>
<tr>
<td>ATTENUATED ERROR CORRECT THEORY</td>
<td>Testing creates a more reliable error signal than studying.</td>
<td>(Mozer et al., 2004)</td>
</tr>
</tbody>
</table>
Conclusion

The effect that testing has on retention for nondeclarative tasks is a significant topic for study and is one that has clear implications for education and general learning. While the testing effect has been frequently studied, however the tasks used in these studies are limited to ones that require explicit, declarative memory strategies to complete. It is currently unclear if the testing effect also extends to tasks that are learned via nondeclarative memory strategies. The present dissertation was designed to extend the testing effect literature by answering some key questions. With regards to nondeclarative memory in particular: 1) Does studying cause immediate memory enhancements to performance? 2) Does testing reduce the amount of forgetting relative to studying? 3) Does the type of learning strategy affect memory retention after testing? 4) Is the testing effect stronger for declarative or nondeclarative memory? The effect of testing on memory is robust and has been shown in several varying settings (Adesope et al., 2017), therefore it is entirely possible that testing enhances retention for nondeclarative memory tasks. The three experiments were designed to answer these questions and further explore the effect of testing on retention.
EXPERIMENT 1

Method

Word Fragment Completion Task

Experiment 1 aimed to understand the effect of testing on retention in a task that requires nondeclarative memory. Participants completed a word fragment completion task by initially rating 40 words based on how positive or negative the participants found the words. Half of the words that were studied were restudied (study) and the other half were tested (test) after the initial study phase. Participants completed the final evaluation phase either immediately after the training phase (immediate condition) or 48hrs later (delay condition). The memory (priming) of the tested and studied words was assessed during this evaluation phase, and a testing effect would be found if there was less forgetting for tested words as compared to studied words after 48 hours (Figure 5). Participants either completed a typical nondeclarative word fragment completion task (nondeclarative memory condition), or a version designed to tap into declarative memory (declarative memory condition). While the testing effect has been shown with single word lists, it has not been shown when tested with word fragments. Single word lists do not have a suitable nondeclarative analog in the way that the word fragment completion task does. It is important to show a testing effect in the declarative condition so that if there is no testing effect for nondeclarative memory, it would not be because of the task used. Since the testing effect has been shown with single word lists, it was predicted that there would be a testing effect in this condition.

It was less clear if there would be a testing effect for nondeclarative memory. If nondeclarative memory responds similarly to declarative memory for testing, then testing would
increase nondeclarative memory by showing less forgetting for the tested words, showing a testing effect for nondeclarative memory. It was also possible that testing would have no benefit as compared to re-studying for nondeclarative memory. If this is true, then the long-term memory for the tested words would not be greater than the studied words in the nondeclarative memory condition.
Figure 5: Procedure for both the immediate condition (top) and the delay condition (bottom).
Method

Participants & Design

Participants (N = 110, undergraduate students from the University of Maine) arrived for a study “Word Ratings”. All participants were required to have normal or corrected vision and received course credit for their participation. All participants were randomly assigned to complete either the nondeclarative memory type condition or the declarative memory type condition as well as an immediate or delay timing condition. All participants completed a study phase, a test phase, and a final evaluation phase either immediately after the test phase or approximately 48 hours after the test phase (depending on if they were in the immediate or delay condition).

Materials

The stimuli used in this experiment were words and corresponding graphemic fragments obtained from Roediger et al., (1992) (see Appendix). Sixty total words were selected, and each word is between 6 and 8 letters long and have been selected due to them having low frequency in use. Low frequency words have been shown to elicit greater priming for the word fragment completion task (Roediger et al., 1992).

---

2 An n = 30 per condition was chosen as a conservative estimate for an N needed to detect an effect. Much of the testing effect literature finds a testing effect with a large effect (for example Roediger et al., 2006 found an effect d = .95). Because the task here is novel with the testing effect, the estimated effect size is smaller (d = .50) than what is often found to err on the side of caution. 10 participants did not finish the final evaluation phase.
Procedure

Nondeclarative Condition

The procedure is shown in Figure 5. After providing informed consent participants began with the study phase. During the study phase, participants were asked to rate 40 words on how positive or negative they think each word is. The scale for rating is as follows: extremely positive, moderately positive, slightly positive, neither positive nor negative, slightly negative, moderately negative, and extremely negative. Immediately after the study phase, participants rated half of the words presented in the initial training phase an additional time. After the study phase, participants were shown incomplete word fragments from the other half of words and were asked to respond by typing in the first word that comes to mind that completes the fragment. For example, a participant may be shown the fragment “d_n_s_a_r”. The participant would type in the word “dinosaur”. The word fragments were words from the initial study phase (the half that were not restudied) and allow only one legitimate completion. If the participants do not complete the word fragment with the correct word, corrective feedback was presented in the form of the correct word.

Declarative Condition

The declarative condition was identical to the nondeclarative condition except for the instructions shown to the participants. In the declarative condition, participants were asked to remember the words learned during the study phase in addition to rating the words. The declarative condition was designed as an analog to research in the testing effect literature, primarily studies using single word lists that show a testing effect (Carpenter & Delosh, 2006; Hogan & Kintsch, 1971; McDaniel & Masson, 1985; Rowland, 2014; Tulving, 1967; Zaromb &
Roediger, 2010). During the practice test phase, they were told to complete the fragments using words they had rated.

**Evaluation Phase**

Participants completed a final evaluation phase either immediately after the test phase (immediate condition), or 48 hours after (delay condition). During the evaluation phase, participants were given incomplete word fragments. A third of the fragments were from words that were restudied during the study phase, another third are from words that were shown as incomplete fragments in the test phase, and the final third were new words not presented to the participants earlier in the experiment. This design makes it possible to study the amount of learning (priming) for words that were tested during the training phase and for words restudied during the training phase. Therefore, the study design was a mixed-subjects design with the independent variables being: 1) type of training for the words (declarative or nondeclarative) as a between-subjects factor, 2) whether the words are studied or tested as a within-subjects factor, and 3) the delay between the test phase and the study phase (immediate or 48 hours later) as a between-subjects factor predicting the amount of priming as the dependent variable.

**Analyses**

The primary dependent variable of interest is the amount of priming for each participant. The mean priming score was calculated by subtracting the number of correctly completed word fragments for the studied and tested words by the number of correctly completed word fragments for the novel words. First, significance for priming was investigated for the word fragments presented during the evaluation phase using a one sample t-test testing if the priming score is significantly greater than 0. It was predicted that there would be significant priming ($p < .05$) for participants in both declarative and nondeclarative conditions, as well as those in the delay and
immediate conditions. Forgetting is shown by a lower mean priming score for the delay condition as compared to the immediate.

It was predicted that in the declarative condition, there would be a testing effect. This means that there would be significantly less forgetting for the tested words compared to the studied words. The predictions for the nondeclarative were not as clear. Since this is the first research to investigate the testing effect, it depends on if testing operates similarly for declarative memory as it does for nondeclarative memory. Previously, several testing effect theories were presented that could be used to argue for a potential testing effect with nondeclarative memory. If these theories apply to nondeclarative memory as well, then a testing effect would be expected for the nondeclarative condition. A testing effect for nondeclarative memory would look similar to one for declarative memory, with decreased forgetting as a result of testing.

Results

Final Evaluation Phase

Declarative

Initial inspection of the data for the participants in the declarative condition (n = 48) suggests that there was greater priming for the tested words as compared to the studied words, as well as greater priming during the immediate condition as compared to the delay condition (Figure 6). Participant’s performance (the amount of priming) for the studied and tested words was calculated by subtracting the amount of correctly completed novel word fragments from the amount of correctly completed studied or tested words during the evaluation phase. A 2 timing (immediate vs. delay) x 2 training (study vs. test) mixed model Analysis of variance (ANOVA) showed a significant main effect of timing [F(1,47) = 116, p = .002, η² = .187] and training [F(1,47) = 116.090, p < .001, η² = .349]. There was no significant interaction [F(1,47) = 0.003, p
These data suggest that the amount of priming drops for both the studied and tested words after 48hrs similarly, with a general overall benefit from testing.

Figure 6: Performance for both the studied and tested words during the final evaluation phase for the declarative memory condition. Performance is measured by subtracting the amount of correctly completed word fragments for the novel words from the studied and t
Nondeclarative

Initial inspection for the data in the nondeclarative condition (n = 59) suggests that the priming for the tested words are greater than the priming for the studied words and that priming drops for the tested and studied words after 48hrs (Figure 7). A 2 (timing) x 2 (training) mixed model ANOVA revealed a significant main effect of training [F(1,58) = 268.058, p < .001, η² = .391] and a significant main effect of timing [F(1,58) = 68.884, p < .001, η² = .543]. These main effects were qualified by a significant interaction between the timing and training condition [F(1,58) = 32.474, p < .001, η² = .047]. In order to investigate the interaction, difference scores were computed by subtracting the priming for the studied words from the priming for the tested words. An independent samples t-test revealed a greater difference in these scores for the immediate condition as compared to the delay condition [t(58) = 5.70, p < .001, d = 1.48]. These results suggest that there was a greater benefit for testing for those that took the final evaluation phase immediately after the test phase, as compared to those who took it 48 hours later.
Figure 7: Performance for both the studied and tested words during the final evaluation phase for the declarative memory condition. Performance is measured by subtracting the amount of correctly completed word fragments for the novel words from the studied and tested words.
Discussion

A testing effect would have been shown in Experiment 1 if there was decreased forgetting for the tested words as compared to the studied words. Results from Experiment 1 do not show an interaction between the timing and training type conditions for the declarative memory condition and an interaction in the nondeclarative condition driven by high priming in the immediate condition for the tested words. For both memory conditions, there was a decrease in priming over time. Because of this, there is no protective benefit to testing in that there was similar forgetting between the studied and tested words. One concern is that the performance for the studied words dropped to zero after 48hrs, creating a floor effect. It is not possible for there to be a greater decrease in performance when performance drops to zero. It should be noted however that there was a general benefit to testing in that there was greater priming for the tested words as compared to the studied words regardless of the timing condition.
Chapter 3

EXPERIMENT 2

Word Fragment Completion Task

Background

One potential limitation presented in Experiment 1 involves showing feedback to the participants during the test phase. Feedback was shown to the participants in the test phase to try and closely match the exposure of the studied and the tested words. Feedback may have served as an instruction to participants in the nondeclarative condition to use the words shown in the study phase, instead of using the first word that comes to mind. This realization would cause the participant to use declarative memory to complete the task even in the nondeclarative memory condition (Howard, 1988). Experiment 2 of this dissertation aimed to replicate the overall benefit of testing for both memory conditions as well as extend the findings by using a procedure in which participants were not shown feedback.

The primary difference between Experiment 1 and Experiment 2 was that in Experiment 2, feedback was not provided to the participants at any point. While feedback is necessary for certain forms of nondeclarative learning, feedback is not necessary for priming (Tulving & Schacter, 1990). Removing feedback in Experiment 2 could affect the results in multiple ways. One, not providing feedback could decrease the amount of priming for the tested words as compared to the studied words. During the testing phase, participants saw a word fragment, and if they correctly completed it, they saw that word and processed it an additional time. If they were unable to complete it correctly, they did not receive feedback, and would not see the word an additional time. With feedback, as in Experiment 1, there is a guarantee that they could see
and process each word a second time during the testing phase. It is likely that reducing the amount of processing for each word would reduce the memory of those words.

Second, feedback itself may have caused participants to shift from using nondeclarative memory to declarative memory in the nondeclarative memory condition. Removing feedback may cause participants to be more likely to use nondeclarative memory. If this is the case, and if participants in Experiment 1 were using declarative memory in the nondeclarative condition, then a different pattern of results would be shown in the nondeclarative condition of Experiment 2. In Experiment 1, tested words were remembered at a greater rate than the studied words in both memory conditions. Since feedback is not necessary for priming, it is possible that removing feedback would not change this and in Experiment 2 there would still be less forgetting for tested words. Experiment 2 aimed to replicate overall benefit of testing and measure a testing effect with the absence of feedback.

The results of Experiment 1 did not show a traditional testing effect in either the declarative or nondeclarative memory conditions. Instead, there was simply an overall benefit of testing in both the declarative and nondeclarative memory conditions. The goal of Experiment 2 was to expand upon Experiment 1, address the issue of feedback within Experiment 1, and to investigate if the results of Experiment 1 would replicate when no feedback is presented. Also, participants were asked how they completed the word fragments in order to better understand which participants used declarative or nondeclarative memory to complete the task. Based on the results of Experiment 1, it was expected that there would be overall greater priming for the tested words as compared to the studied words. It was also expected that there would be a decrease in priming from the immediate to the delay condition for both training types.
Method

Participants & Design

Participants were recruited from both Amazon’s Mechanical Turk online application, and from the University of Maine’s undergraduate population. Amazon’s Mechanical Turk allows for researchers to post surveys and studies for individuals for pay. Participant’s that completed Experiment 2 within Amazon’s Mechanical Turk were compensated with $4.00, native English speakers, residents of the United States, and above the age of 18.

Table 2: Number of participants in each of the conditions, as recruited from either Amazon’s Mechanical Turk or the University of Maine.

<table>
<thead>
<tr>
<th></th>
<th>AMAZON’S MECHANICAL TURK</th>
<th>UNIVERSITY OF MAINE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DECLARATIVE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate</td>
<td>30</td>
<td>9</td>
</tr>
<tr>
<td>Delay</td>
<td>6</td>
<td>26</td>
</tr>
<tr>
<td><strong>NONDECLARATIVE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate</td>
<td>31</td>
<td>17</td>
</tr>
<tr>
<td>Delay</td>
<td>4</td>
<td>25</td>
</tr>
</tbody>
</table>
Participants (n = 77, undergraduate students from the University of Maine\(^3\) and n = 71 from Amazon’s Mechanical Turk; Table 2) completed a study “Word Ratings”. The samples from the University of Maine and from Amazon’s Mechanical Turk were combined for all analyses. All participants who completed Experiment 2 from the University of Maine were required to have normal or corrected vision and received course credit for their participation. All participants were randomly assigned to complete either a nondeclarative memory type condition or a declarative condition as well as an immediate or delay timing condition. All participants completed a study phase, a test phase, and a final evaluation phase either immediately after the test phase (immediate condition) or 48 hours after the test phase (delay condition).

**Materials**

The stimuli used in this experiment were the same words used in Experiment 1 and were the words and corresponding graphemic fragments obtained from Roediger et al., (1992) (see Appendix A).

**Procedure**

**Nondeclarative Condition**

The procedure was identical to Experiment 1 except feedback was not provided during the testing period (Figure 8). Also, participants were asked how they completed the word fragments at the end of the experiment.

\(^3\) An n = 30 per condition was chosen as a conservative estimate for an N needed to detect an effect. Much of the testing effect literature finds a testing effect with a large effect (for example Roediger et al., 2006 found an effect d = .95). Because the task here is novel with the testing effect, the estimated effect size is smaller (d = .50) than what is often found to err on the side of caution. More than the target 30 participants were collected in each condition to try and balance the conditions for recruitment from the University of Maine and Amazon’s Mechanical Turk.
**Declarative Condition**

The procedure was identical to Experiment 1 except feedback was not be provided during the testing period (Figure 8). Also, participants were asked how they completed the word fragments at the end of the experiment.
Figure 8: Procedure for both the immediate condition (top) and the delay condition (bottom).
Analyses and Results

The data was analyzed in the same way as the data in Experiment 1. The mean priming score was calculated by subtracting the number of correctly completed word fragments for the studied and tested words by the number of correctly completed word fragments for the novel words. Forgetting is shown by a lower mean priming score for the delay condition as compared to the immediate. Within both the declarative and nondeclarative memory conditions, the priming scores for the tested and the studied words were compared across timing conditions using a mixed model ANOVA. The training type of word, tested or studied, is a within-subjects factor. The timing condition, immediate or delay, is a between-subjects factor.

Results

Final Evaluation Phase

Declarative

Initial inspection of the data for the participants in the declarative condition (n = 71) suggests a difference in performance between the studied and tested words (Figure 9). The data also suggests an interaction between the training condition and the study condition. Similar to Experiment 1, the performance (as measured by the amount of priming) for the studied and tested words was calculated by subtracting the amount of correctly completed novel word fragments from the amount of correctly completed studied or tested words during the evaluation phase.

A 2 (timing) x 2 (training) mixed-model ANOVA revealed a significant main effect of training type [F(1,69) = 65.146, p < .001, η² = .186] and no significant main effect of timing condition [F(1,69) = .038, p = .846, η² = .001]. The main effect of training was qualified by a
significant interaction \( [F(1,69) = 48.276, p = .004, \eta^2 = .025] \). As in Experiment 1, difference scores were computed by subtracting the priming for the studied words from the priming for the tested words. An independent samples t-test revealed a significant difference in these scores between the immediate and the delay conditions \( [t(69) = 2.98, p = .004, d = 0.711] \). These results suggest that there was a greater benefit of testing for those that took the final evaluation phase 48hrs after the test phase, as compared to those who took it immediately. While this pattern of results is not typical within the testing effect literature, due to a lack of forgetting, a testing effect is still shown due to greater priming memory after a delay as compared to immediately after.
Figure 9: Performance for both the studied and tested words during the final evaluation phase for the declarative memory condition. Performance is measured by subtracting the amount of correctly completed word fragments for the novel words from the studied and tested words.
NonDeclarative

Initial inspection of the data in the nondeclarative condition (n = 76) shows a similar pattern to the declarative condition (Figure 10) in that the priming for the tested words were greater than the priming for the studied words. A 2 (timing) x 2 (training) mixed model ANOVA revealed a significant main effect of training type \([F(1,75) = 136.278, p < .001, \eta^2 = .235]\) and no significant effect of timing \([F(1,75) = 1.065, p = 0.305, \eta^2 = .014]\). The main effect of training is qualified by a significant interaction \([F(1,75) = 13.386, p < .001, \eta^2 = .151]\). Difference scores were computed by subtracting the priming for the studied words from the priming for the tested words. An independent samples t-test revealed a significant difference in these scores between the immediate and the delay conditions \([t(75) = 3.659, p < .001, d = 0.860]\). These results suggest that there was a greater benefit for testing for those that took the final evaluation phase 48hrs after the test phase, as compared to those who took it immediately. Similar to the declarative condition, this shows a testing effect, due to greater priming memory after a delay as compared to immediately after.
Figure 10: Performance for both the studied and tested words during the final evaluation phase for the nondeclarative memory condition. Performance is measured by subtracting the amount of correctly completed word fragments for the novel words from the studied and tested words.
Task Strategy

An important aspect to consider when using tasks thought to require primarily declarative or nondeclarative memory is whether or not participants are actually using the intended strategy. With a task like the typical nondeclarative version of the word fragment completion task, it is possible that participants become test aware. A test aware participant is someone who realizes that the word fragments they are completing are from words they had seen earlier in the experiment. Test aware participants use declarative memory strategies instead of the intended nondeclarative strategies (Howard, 1988). Previous research has shown that with a similar task, the word stem completion task, close to half of the participants in the nondeclarative condition become test aware (Bowers & Schacter, 1990).

At the end of Experiment 2, participants were asked to write down anything they did to help them complete the word fragments. This was used to determine whether a participant was test aware. This was only done for the participants in the nondeclarative, since by design the participants in the declarative condition were made test aware. Participants who indicated or referred to the words they had seen previously are considered test aware. For example, a participant wrote “the only reason I got some of the words was because they were the ones that stood out to me in the first part of the survey about ranking the positivity of the words”. From this it was clear that they were test aware. The following analyses were focused on dissecting the affect that being test aware has on both the overall priming and the testing effect.

First, the proportion of test aware participants are calculated for participants in the nondeclarative condition. In the delay condition, 55% of participants were test unaware, meaning that they did not indicate that they were aware that at least some of the word fragments during the evaluation phase came from the initial study phase. In the immediate condition, 58% of
participants were test unaware. The timing condition did not have any impact on whether the participants were test aware.

Initial inspection of the data for participants in the nondeclarative condition that were test aware suggests an overall benefit of testing (Figure 11). A two-way, mixed-model ANOVA, with the timing condition (immediate vs delay) and the training condition (study vs tested words) showed both a main effect of timing \([F(1,31) = 4.316, p = .046, \eta^2 = .122]\) and a main effect of training condition \([F(1, 31) = 88.444, p < .001, \eta^2 = .251]\). These main effects were not qualified by an interaction \([F(1,31) = 1.816, p = 1.819, \eta^2 = .005]\). These results suggest that for those aware of the test manipulation, that there was more priming for tested words and overall, there were equal amounts of forgetting for the tested and studied words.

Initial inspection of the data for those who were test unaware suggests an overall benefit of testing as compared to studying and less forgetting for the tested words (Figure 12). A two-way, mixed-model ANOVA, with the timing condition (immediate vs delay) and the training condition (study vs tested words) showed a main effect for the testing condition \([F(1, 42) = 61.920, p < .001, \eta^2 = .226]\) and no main effect of the timing condition \([F(1, 42) = 0.299, p = .587, \eta^2 = .007]\). The main effect of testing was qualified by an interaction between the timing and testing condition \([F(1, 42) = 52.025 , p < .001, \eta^2 = .046]\). Difference scores were computed by subtracting the priming for the studied words from the priming for the tested words. An independent samples t-test revealed a significant difference in these scores between the immediate and the delay conditions \([t(42) = 3.566, p = .001, d = 1.118]\). These results suggest that there was a greater benefit for testing for those that took the final evaluation phase 48hrs after the test phase, as compared to those who took it immediately. Therefore, a testing effect is shown due to greater priming memory after a delay as compared to immediately after.
Figure 11: Performance for both the studied and tested words during the final evaluation phase for the test aware participants in the nondeclarative memory condition. Performance is measured by subtracting the amount of correctly completed word fragments for the novel words from the studied and tested words.
Figure 12: Performance for both the studied and tested words during the final evaluation phase for the test unaware participants in the nondeclarative memory condition. Performance is measured by subtracting the amount of correctly completed word fragments for the novel words from the studied and tested words.

**Discussion**

A typical testing effect, as illustrated by previous work, would have been shown if the amount remembered (measured by priming) decreased less for the tested words than it did for the studied words. In the declarative condition, priming decreased for the studied words and did not decrease for the tested words over a 2-day period. While this does not show a typical pattern of
results where testing would still result in forgetting, only less forgetting relative to studying, this is still a testing effect. Words that were tested, instead of studied, were remembered at a greater rate. In the nondeclarative condition, a similar trend was shown, greater remembering for the tested words as compared to the studied words. This shows a testing effect with nondeclarative memory.

Participants were recruited from both the University of Maine and Amazon’s Mechanical Turk (Table 2). Participants from both samples were at least 18 years of age but there are some key differences to consider. For one, students from the University of Maine sample are all currently enrolled in college and completed this task to get credit for a college course. Participants from the Amazon’s Mechanical Turk sample are distributed throughout the United States. Despite these differences, there were no differences in performance found between the samples. For example, there was no difference between the samples for either the studied (p = 0.159) or the tested words (p = 0.138) within the immediate, Nondeclarative condition. Similarly, there was no difference found between the two samples for any other conditions. It should also be noted that there were large differences in the sample sizes between the Amazon’s Mechanical Turk sample and the University of Maine sample in each condition. Despite the difference in sample size, there were no differences between the two samples and the two samples were combined for analyses.

The goal of Experiment 2 was to understand the effect that testing has on long term memory and if there is a testing effect for nondeclarative memory. Not only was a testing effect shown in the declarative and nondeclarative conditions, but it was shown for those that were test unaware. For those that were test unaware, performance tended downwards for the studied words and trended upwards for the tested words. For those that were test aware, performance
decreased for the studied and tested words. If participants who are aware are using declarative memory and those that are unaware are using nondeclarative strategies, then these results show a stronger testing effect for nondeclarative memory. The results of Experiment 2 contrast the results from Experiment 1. In Experiment 1, there was a greater benefit of testing for the immediate as compared to the delay condition for both the declarative and nondeclarative memory condition. Overall, the results for Experiment 2 show evidence for a testing effect with nondeclarative memory.
Chapter 4

EXPERIMENT 3

Weather Prediction Task

Background

The goal of Experiment 3 was to extend the results of Experiment 1 and 2 using a different task that is thought to tap into a different type of nondeclarative memory. Where Experiment 1 and 2 showed a testing effect using a word fragment completion task (a priming task), Experiment 3 investigates the testing effect using the weather prediction task. Experiment 3 utilized a 2 (timing) x 2 (training) between-subjects design. Participants were trained to learn the associations between geometric cues and weather patterns (see Knowlton et al., 1994; Poldrack et al., 2001, for examples of other studies using the same task).

The weather prediction task is different than the word fragment completion task in several key areas. First, the weather prediction task is thought to require a different kind of nondeclarative memory, procedural memory (Knowlton et al., 1996, 1994), whereas the word fragment completion task requires perceptual priming (Roediger & McDermott, 1993). A perceptual priming test can challenge the perceptual system by presenting words in a fragmented form and does not require feedback to learn. Procedural learning is marked by gradual changes in performance due to repeated experience. It is possible that testing could function differently for procedural memory than it does for perceptual priming.

Another area in which the weather prediction task differs from the word fragment completion task is that it is probabilistic. Each geometric cue is associated with either the rain or the sun with a different probability, none of which are deterministic (0 or 100%). The
probabilistic nature of this task is thought to encourage a nondeclarative, procedural, learning strategy (Knowlton et al., 1994).

Different variations of the weather prediction task have been shown to recruit either more declarative or nondeclarative memory. Poldrack and colleagues (2001) used a feedback and a paired-associates version of the weather prediction task. The paired-associates version of the tasks has subjects view the geometric card patterns and the correct weather pattern simultaneously. The subjects learn the association between the geometric cues and the weather patterns by explicitly memorizing which weather patterns are associated with which cues. This is similar to the word-pair association tasks seen in the testing effect literature where subjects learn the associations between two words in a pair (ex. Allen, Mahler, & Estes, 1969; Carpenter, 2009; Carpenter et al., 2006; Carrier & Pashler, 1992; Kornell, Bjork, & Garcia, 2011; Pyc & Rawson, 2010; Toppino & Cohen, 2009). This version has been shown to recruit the neural systems associated with declarative memory (Poldrack et al., 2001). Therefore, it is expected that this version of the weather prediction task could benefit from testing in a comparable way that the declarative tasks in the testing effect literature do.

The other version of the weather prediction task is a feedback version that emphasizes learning via trial and error. This version has subjects view the geometric shapes and make decisions about which weather pattern these geometric shapes predict. For each response, the subjects are presented with feedback in the form of the correct weather pattern and a face, either smiling or frowning. This version of the task is thought to require nondeclarative learning. This has been shown to activate neural systems associated with non-declarative learning (Poldrack et al., 2001).
There were multiple predictions made for Experiment 3. Based upon Experiments 1 and 2 which showed a decrease in memory from the immediate to the delay time points, it was predicted that there would be a decrease in accuracy for both the participants in the study and the test condition. The training factor had two conditions, a study and test condition. Based on previous research showing that the study condition is associated with declarative memory (specifically for the weather prediction task), and the test condition is associated with nondeclarative memory (Poldrack et al., 2001), it was predicted that participants in the study condition would use strategies consistent with declarative memory to complete the task and those in the test condition would have a higher proportion of participants using a nondeclarative memory strategy as compared to the study condition (Gluck et al., 2002). In terms of a testing effect, it was predicted based on the results of Experiment 1 and 2 that there would be greater accuracy for those in the testing condition, with no difference in accuracy between the testing and the study conditions. Not only does Experiment 3 expand on the findings of Experiments 1 and 2, which are the first studies investigating the testing effect with a nondeclarative memory task, but it is also the first to use a probabilistic learning task as well.

**Method**

**Participants & Design**

Participants (N = 122, undergraduate students from the University of Maine\(^4\)) arrived for a study “Perceptual and Cognitive Memory”. All participants were required to have normal or

---

\(^4\) An n = 30 per condition was chosen as a conservative estimate for an N needed to detect an effect. Much of the testing effect literature finds a testing effect with a large effect (for example Roediger et al., 2006 found an effect d = .95). Because the task here is novel with the testing effect, the estimated effect size (d = .50) is smaller than what is often found to err on the side of caution. Because of the low amount of participants using a declarative strategy additional participants were recruited.
corrected vision and received course credit for their participation. Participants were randomly assigned to either a testing or a study condition as well as an immediate or a delay condition. Participants in the testing condition completed a feedback version of the weather prediction task and participants in the study condition completed a paired-associates version of the weather prediction task. Participants in the immediate condition completed a final evaluation phase immediately after the learning phase of the study and those in the delay condition completed the final evaluation phase 48hrs after the initial learning phase.

**Materials**

**Feedback Weather Prediction Task**

![Weather patterns used in Experiment 3](image)

*Figure 13: Weather patterns used in Experiment 3*

![Geometric cues used in Experiment 3](image)

*Figure 14: Geometric cues used in Experiment 3*
Participants in the testing condition completed a feedback version of the weather prediction task (Knowlton et al., 1994). In this task, participants are asked to learn which of two outcomes (rain or sun) is predicted by the combination of one, two, or three different cues (Figure 13 and Figure 14). Each cue is independently associated to an outcome (see Table 3). One, two, or three cues can appear on the screen during each trial, for a total of 14 patterns. The outcome for each trial is calculated according to the probabilities of the outcome and the cards occurring together (see Table 3). The subjects completed a total of 50 trials per block for 4 blocks.
Table 3. Total frequency of each pattern, the number of times each pattern occurred, and the outcomes for Experiment 3 (Gluck et al., 2002).

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Cards present</th>
<th>Sun</th>
<th>Rain</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0001</td>
<td>17</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>B</td>
<td>0010</td>
<td>7</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>C</td>
<td>0011</td>
<td>24</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td>D</td>
<td>0100</td>
<td>2</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>E</td>
<td>0101</td>
<td>10</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>F</td>
<td>0110</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>G</td>
<td>0111</td>
<td>17</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>H</td>
<td>1000</td>
<td>2</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>I</td>
<td>1001</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>J</td>
<td>1010</td>
<td>2</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>K</td>
<td>1011</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>L</td>
<td>1100</td>
<td>2</td>
<td>24</td>
<td>26</td>
</tr>
<tr>
<td>M</td>
<td>1101</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>N</td>
<td>1110</td>
<td>2</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
During each trial, one, two, or three geometric cues appeared on the screen. Participants were asked to select which outcome is associated with the presented cards by using the mouse to click on the correct outcome, either rain or sun. After responding, feedback was provided. After a correct response, the word “CORRECT” appeared in green and was accompanied by a 1 s, 500 Hz tone, and the correct outcome. After an incorrect response, the word “WRONG” appeared in red and is accompanied by a 1 s, 200 Hz tone and the correct outcome. After which, the screen was blanked for 500 ms prior to the appearance of the next pattern. Summary feedback in the form of percent correct was presented at the bottom of the screen during each trial and at the end of each block.

**Paired-Associates Weather Prediction Task**

The task was identical to the feedback version of the weather prediction task except that they were shown the correct weather response. Participants were asked to use the presented correct weather pattern during each trial to learn the relationship between the cards and the weather and were informed that they would be tested on these relationships.

**Procedure**

After providing informed consent, participants were randomly assigned to one of the timing conditions (immediate or delay) and to one of the study conditions (test or study). Participants completed either the feedback version of the weather prediction task, if they were in the test condition, or the paired-associates, if they were in the study condition. Participants in the immediate condition completed the test phase immediately after training and participants in the delay condition returned after 48 hours to complete the final evaluation phase. The final evaluation phase was identical to the feedback version of the weather prediction task except no feedback was presented.
Strategy Analyses

A crucial factor to consider is how participants approach the task. Because the cue-outcome associations are probabilistic, it is often thought that subjects learn these associations incrementally (Gluck & Bower, 1988). This is not always the case, in fact there are a number of different strategies that subjects could use to solve the weather prediction task and it is possible that different memory systems could underlie strategies (Gluck, Shohamy, & Myers, 2002). These strategies could highlight differences in learning that simply the percent correct would not show.

There are three primary strategies participants can use to solve the weather prediction task. The first is the one-cue learning strategy. Subjects that adopt this strategy base their responses on the presence or absence of a single cue. For example, one may always predict the sun whenever they see the oval card. The second is the multi-cue learning strategy. Subjects that adopt this strategy base their responses on the combination of cues presented. For example, they may choose the sun pattern whenever they see both the triangle card and the diamond card. The third is the singleton learning strategy. Subjects who adopt this strategy learn the correct response to singleton patterns (patterns when only one card appears) and guess on the remaining trails. For example, when they see only the triangle or only the diamond card, they may choose the rain pattern and if they see only the square or only the oval card, they may choose the sun pattern.

The percent optimal, the performance by a subject following the strategy, varies depending on which strategy is used. The percent optimal for the strategies is as follows: multi-cue, 100%; singleton, 75%; one-cue using the highly-predictive cues, 87.5%; one-cue using the less-predictive cues, 66%. Therefore, it is generally advantageous to use a multi-cue strategy.
verses a single cue strategy. It is assumed that the weather prediction task is solved via incremental, non-declarative memory (Knowlton et al., 1994, 1996). This could depend on the strategy used, as the one-cue strategy could be learned in a way that is easily verbalizable (Gluck et al., 2002). For example, the strategy “Choose the “rain” when the triangle is present” could result in performance significantly higher than chance and is a strategy that requires declarative memory. Therefore, it is possible for subjects to complete the task using strategies that theoretically would recruit either declarative or nondeclarative memory processes.

Because this task can be completed using strategies that theoretically recruit either declarative or nondeclarative memory, the weather prediction task can be used to investigate the difference in the magnitude of the testing effect for the different memory systems. The paired-associates version of the weather prediction task should encourage declarative memory strategies whereas the feedback version should encourage nondeclarative memory strategies (Poldrack et al., 2001). Therefore, it is expected that in the paired-associates version, a greater proportion of participants would be using either the singleton or a one-cue strategies as compared to the feedback-version. The proportion of participants using a nondeclarative memory strategy for both the feedback and the paired-associates version of the tasks would be measured.

Results

Training Phase

117 participants completed the task (n = 72 for the delay condition and n = 50 for the immediate condition). Analyses were conducted on the percent optimal responses within each block to measure if there was significant learning of the weather pattern/cue relationships. A participant’s percent optimal responses is the percent of their responses that are the optimal response. For example, for pattern A (Table 1), the optimal response is “Sun”. The percent
optimal responses during training phase is only meaningful in the test condition since subjects in the study condition were shown the correct response. The training data was collapsed across the timing conditions since participants in both timing conditions completed the training during the first day. Only training data for participants in the testing condition (n= 63) were useable since there is no measure of performance during training for the study condition. A dependent measures t-test revealed that there was a significant increase in accuracy from block 1 to block 4 \[t(63) = 8.98, p < .001, d = 1.15\]. This shows participants were able to increase their understanding of the pattern/cue relationships from block 1 to block 4 of training within the testing condition.

**Test Phase**

Initial inspection of the data suggests there is no difference in accuracy between the immediate and delay time points for either the study or test condition (Table 4). Participants who completed the task by guessing were not included in these analyses. percent optimal responses during the test phase is compared to investigate whether there is a testing effect. A 2 (timing) x 2 (training) between subjects ANOVA revealed no significant interaction \[F(1,69) = .001, p = .994, \eta^2 = .000\], no significant effect of test \[F(1,69) = 2.057, p = .156, \eta^2 = .028\], and no significant effect of timing \[F(1, 69) = 3.422, p = .069, \eta^2 = .049\] (See Figure 15). These data do not suggest that the training type, whether the pattern/cue relationship was learned via studying or testing, had an impact on how well the relationship was learned. This is consistent with previous literature that used similar methodology (Poldrack et al., 2001). More importantly, these data suggest that there is no decrease in the memory of the pattern/cue relationships after 48hrs regardless of training type. It is possible that either training type is effective for the learning of these relationships.
Table 4: Descriptive statistics for Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Study</th>
<th></th>
<th>Test</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Immediate</td>
<td>Delay</td>
<td>Immediate</td>
<td>Delay</td>
</tr>
<tr>
<td>N</td>
<td>24</td>
<td>37</td>
<td>26</td>
<td>35</td>
</tr>
<tr>
<td>Avg Percent</td>
<td>0.67</td>
<td>0.70</td>
<td>0.69</td>
<td>0.73</td>
</tr>
<tr>
<td>Optimal Responses</td>
<td>0.18</td>
<td>0.16</td>
<td>0.18</td>
<td>0.17</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>0.18</td>
<td>0.16</td>
<td>0.18</td>
<td>0.17</td>
</tr>
</tbody>
</table>
Figure 15: Percent Optimal during the testing phase, split by timing and test conditions.

**Strategy Analyses**

**Strategy Counts**

Participants in all conditions were best fit by models assuming the following strategies, multi-cue, singleton, one-cue using the highly predictive cues, one-cue using the less-predictive
cues, and a guessing model (See Gluck et al., 2002 and Appendix B). These models were fit to the data during the final evaluation phase for both the study and test conditions. The strategy types were split into two categories, declarative strategies and nondeclarative strategies (Error! Reference source not found.). The nondeclarative strategy, the multi-cue strategy, relies on participants combining and using the information from all of the cues presented during each trial and is associated with dorsal striatal activation (Schwabe, 2016). The declarative strategies involve participants choosing the weather pattern generally by making their decisions based upon a single cue. These include the singleton, one-cue, and the guessing-model. These declarative strategies are associated with hippocampal activation. The hippocampus has been shown to be important for declarative learning (Packard et al., 1989).

A chi-square test of independence was performed to examine if strategy varied across training and timing conditions. The relationship between the model used and training condition was not significant $[X^2 (2, N = 119) = 3.527, p = .474]$. The relationship between the model used and timing condition was not also significant $[X^2 (2, N = 119) = 3.527, p = .087]$. The strategy used did not vary across the training conditions (Study vs Test) and it did not vary across the timing conditions (Immediate vs Delay).
Table 5: Descriptive statistics for participants fit by either a nondeclarative, declarative, or guessing model. Included are mean accuracy and the percent of participants who adopted the strategy.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Study Immediate</th>
<th>Study Delay</th>
<th>Testing Immediate</th>
<th>Testing Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.</td>
<td>%</td>
<td>Mean</td>
</tr>
<tr>
<td>Nondeclarative</td>
<td>.83</td>
<td>.05</td>
<td>52%</td>
<td>.84</td>
</tr>
<tr>
<td>Declarative</td>
<td>.07</td>
<td>.13</td>
<td>24%</td>
<td>.44</td>
</tr>
<tr>
<td>Guessing</td>
<td>.05</td>
<td>.07</td>
<td>24%</td>
<td>.60</td>
</tr>
</tbody>
</table>

**Nondeclarative**

Participants who used a nondeclarative strategy percent optimal responses during the test phase are compared to investigate whether there is a testing effect specifically for nondeclarative
memory (Figure 16). A 2 (timing) x 2 (training) between subjects ANOVA revealed no significant interaction [F(1,55) = 1.88, p = .176, \( \eta^2 = .01 \)], no significant effect of test [F(1,55) = 2.18, p = .145, \( \eta^2 = .04 \)], and no significant effect of timing [F(1, 55) = .574, p = .452, \( \eta^2 = .03 \)]. These data do not suggest that testing does differentially affects long term nondeclarative memory depending on training condition.
Figure 16: Percent Optimal during the testing phase, split by timing and test conditions, for those using a nondeclarative strategy.
Discussion

Data from Experiment 3, the weather prediction task, do not give evidence for a traditional testing effect. A traditional testing effect is defined as a small or modest benefit of studying relative to testing when a final evaluation phase is completed immediately after studying (Roediger & Karpicke, 2006b). When the final evaluation phase is delayed, a benefit of testing appears relative to studying. Specifically, less of the tested information is forgotten compared to the re-studied information. A less strict definition may be used where there only needs to be a benefit of testing as compared to studying, but these results do not suggest a testing effect even by this definition.

When considering all participants, those who used declarative and nondeclarative memory strategies to complete the weather prediction task, no difference was found in accuracy for those that learned via studying and those that learned via testing. In addition, no difference in performance was detected between the immediate time point and the time point 48hrs later. This same trend was also found when only looking at participants who used a nondeclarative, multi-cue strategy. Therefore, no testing effect was found. There was no initial benefit to studying and there was no long-term benefit to testing. The question is then, why was no testing effect found for the weather prediction task?

One possibility is that the delay between the immediate condition and the delay condition was not long enough. The testing effect has been shown with delays of only 48 hrs, but with these experiments, performance in both the test and study conditions decreases after 48 hrs (Roediger & Karpicke, 2006a). Performance in Experiment 3 did not decrease for either
condition. It is impossible to detect a testing effect if performance does not change over time. Research has showed that as the delay increases, the strength of the testing effect does as well. Increasing the delay length for the weather prediction task could be important for finding a testing effect. Where accuracy did not decrease after 48hrs, it is possible that with a greater delay of a week or more, performance would decrease. A decrease in accuracy between the immediate and delay conditions would show a testing effect if the decrease was greater for the study condition.

Surprisingly, there was no evidence for a difference in strategy used by participants between the training conditions. Poldrack and colleagues (2001) used a similar weather prediction task paradigm with a feedback and a paired-associates condition and found greater activation of the hippocampus in the paired-associates and greater striatal activation for those in the feedback condition, though they did not fit the participants data to strategy models. The hippocampus has been implicated in declarative memory (Packard et al., 1989), therefore it is reasonable to expect that strategies thought to require more declarative memory (singleton or one-cue) recruit the hippocampus more. In fact, increased hippocampal activation has shown to be associated with the use of the declarative strategies (Schwabe & Wolf, 2012). Because Poldrack and colleagues (2001) found that participants in the paired-associates condition had greater hippocampal activation, it is reasonable to expect that participants in this similar experiment would have greater hippocampal activation and therefore a greater reliance on declarative strategies. This was not found. The training had no significant impact on the model used.

One key difference however between the experiment presented here, and the ones shown in both Poldrack and colleagues, 2001, and Schwabe, 2012, is that the participants decision
strategies in Experiment 3 were fit during the test block during the final evaluation phase, not the training block. The reason for this is because there is no data for the participants in the study condition to fit in the training phase. It is possible that early in learning, during the training phase, participants in the study condition are using declarative strategies at a greater rate than participants in the test condition. Later in training, or during the test phase, participants could shift to a nondeclarative strategy. Overall, it is impossible to determine which strategies participants in the study phase use during training. During test, participants in both study and test conditions were equally likely to use a declarative or nondeclarative strategy.
Chapter 5

General discussion

The findings in the present dissertation contribute to the testing effect literature by being the first research to investigate how testing impacts long term nondeclarative memory retention. Experiment 1 and 2 used the word fragment completion task to investigate how being tested on primed words effects the long-term retention of the priming as compared to re-studied words. Experiment 3 used the weather prediction task to investigate how testing impacts procedurally learned information as compared to studying. The results from Experiment 1 do not show evidence for a testing effect in either the declarative or the nondeclarative memory conditions. Instead, the results only show an overall benefit of testing. While the results from Experiment 2 do not show a classic testing effect, in that testing eliminated forgetting of the primed words, testing still increased retention. Therefore, it is argued that a testing effect for Experiment 2. The words that were tested were forgotten at a lesser rate than those that were studied. The results from Experiment 3 do not support a testing effect for procedural, nondeclarative memory. Taken together results from the three experiments support an overall benefit for testing and a testing effect for nondeclarative memory. The differential results from Experiments 1 and 2, and 3 suggests that the testing effect may only occur within specific types of nondeclarative memory.

Summary

Experiment 1

Testing has been shown to enhance long term declarative memory relative to studying (Roediger & Karpicke, 2006a). Therefore, a prediction for a testing effect for the declarative memory condition was based on previous research showing a testing effect with word lists. As predicted, priming for tested words was greater than for studied words. For the nondeclarative
condition, there was a greater benefit of testing for those that took the final evaluation phase immediately after the test phase, as compared to those who took it 48 hours later. There was no decrease in forgetting as a result of testing. While this initially seems like a lack of a testing effect, the priming for studied words after 2 days was at 0. It is possible that this floor made it impossible to detect a testing effect. Experiment 1 was the first study to investigate how testing impacts long term nondeclarative memory as compared to studying. The results show an overall benefit of testing. In the declarative condition, there was similar rates of forgetting for the studied and the tested words. Overall, Experiment 1 shows a benefit of testing in both declarative and nondeclarative memory in that more information is initially learned for the tested words with a lesser amount of forgetting after two days. However, a traditional testing effect was not shown for either the declarative or nondeclarative memory condition.

**Experiment 2**

Experiment 2 was designed to further assess the impact that testing has on nondeclarative memory (priming specifically). To better ensure that participants in the nondeclarative condition completed the word fragments using nondeclarative memory, feedback was not given during the test phase or the final evaluation phase. In Experiment 1, it is possible that feedback caused participants to associate the word fragments with the words from the study phase. Experiment 2 also differentiated those who used declarative and those who used nondeclarative strategies by classifying participants as test aware and test unaware.

The results from Experiment 2 show a testing effect for the declarative condition, which was predicted given that the testing effect has been shown with tasks that require declarative memory. There was a greater benefit of testing for participants in the delay condition as compared to the immediate condition. The differences between a declarative word fragment
completion task and a word pair memory task used in many testing effect studies are that the
word fragments act as cues and participants rated the words in the study phase. More
interestingly, a testing effect was found for the nondeclarative condition, even when considering
only those who were test unaware. In fact, testing prevented forgetting for the tested words. This
is the first research showing a testing effect for nondeclarative memory. Showing a testing effect
for those that were test unaware further strengthened the argument for a testing effect for
nondeclarative memory.

Experiment 3

Experiment 3 was designed as both a replication of the findings of Experiment 1 and 2, and as an extension using a different type of nondeclarative memory. The task in Experiment 3, the weather prediction task, is thought to be a procedural learning task (Knowlton et al., 1996, 1994) whereas the word fragment completion task is a priming task (Roediger et al., 1992). The weather prediction task could be completed using various strategies, some of which rely on nondeclarative memory, and others rely on declarative memory. 51% of the participants in both the study and test conditions adopted an optimal, nondeclarative strategy. 27% of participants guessed and performed near chance and the remaining 22% adopted a suboptimal, declarative strategy and performed near chance. This contrasts with previous research showing a majority of participants using declarative strategies and still performing above chance (Gluck et al., 2002).

The primary question for Experiment 3 was if there is a benefit to testing for procedural, nondeclarative memory like there is for priming nondeclarative memory, as shown in Experiment 2. When considering all participants, there is no difference between those who learned by studying the weather pattern associations, and those who learned by testing. Because those who used a declarative strategy performed near chance, and the primary focus of this
dissertation is on nondeclarative memory, it is important to focus just on the participants who used a nondeclarative strategy. There was no difference in accuracy between those who learned by studying or testing and there was no difference in accuracy between the immediate or the delay time points. These results show that the procedural memory learned is more resistant to forgetting and it showed no evidence for a testing effect with procedural memory.

**Testing effect and declarative memory**

The testing effect has been shown countless times with research designs that involve participants learning something to consciously remember it for a later date, all examples of declarative memory (Roediger & Karpicke, 2006b). It has been shown with free-recall tests (e.g. Hogan & Kintsch, 1971; Lachman & Laughery, 1968; Roediger & Karpicke, 2006a; Tulving, 1967; Wheeler & Roediger, 1992), cued-recall (e.g. Carpenter, Pashler, & Vul, 2006), prose passage (e.g. Chan et al., 2006; Glover, 1989; Roediger & Karpicke, 2006a; Spitzer, 1939), single word lists (Carpenter & Delosh, 2006; Hogan & Kintsch, 1971; McDaniel & Masson, 1985; Rowland, 2014; Endel Tulving, 1967; Zaromb & Roediger, 2010), word-pairs (Allen et al., 1969; Carpenter, 2009; Carpenter et al., 2006; Carrier & Pashler, 1992; Kornell et al., 2011; Pyc & Rawson, 2010; Toppino & Cohen, 2009) and not word lists or verbal materials (for example see Carpenter & Pashler, 2007; Kang, 2010; Wheeler & Roediger, 1992).

Each experiment in this dissertation had a condition that was designed to measure a testing effect using tasks aimed to specifically recruit declarative memory. In Experiment 1 and 2, participants in the declarative condition of the word fragment completion task viewed a list of words, one at a time, to be remembered later. This is similar to the research showing a testing effect using single-word lists (Carpenter & Delosh, 2006; Hogan & Kintsch, 1971; McDaniel & Masson, 1985; Rowland, 2014; Endel Tulving, 1967; Zaromb & Roediger, 2010). Then
participants were tasked to complete word fragments with the words that they had seen previously, similar to the research using cued-recall (e.g. Carpenter, Pashler, & Vul, 2006).

Because of the similarities to previous research showing a testing effect, it was predicted that there would be a testing effect for the declarative conditions in Experiments 1 and 2. In Experiment 1, there was not a traditional testing effect. Memory for the words was greater for the tested words as compared to the studied words, but the rates of forgetting were not different. In Experiment 2, a more traditional testing effect was shown. Memory was overall greater for the tested words, and there was less of a decrease in memory over time for the tested words.

In Experiment 3, participants in the declarative condition completed the weather prediction task by studying the weather-card associations one at a time. While this is most similar to the research showing a testing effect with nonverbal materials (Carpenter & Pashler, 2007; Kang, 2010; Wheeler & Roediger, 1992), the experimental design was not similar enough as to make a strong prediction simply based on that. The prediction that there would be a testing effect came from previous research showing that the declarative version of the weather prediction task caused participants to more frequently use strategies that require declarative memory (Schawbe & Wolfe, 2012) and the testing effect is a robust finding with tasks that require declarative memory. There was no testing effect with the either the declarative version of the weather prediction task, or with participants that used a declarative strategy. This is because participants who used declarative strategies to complete the weather prediction task performed near chance during the final test phase. Because of this, it would be impossible to detect a testing effect. Therefore, it is inconclusive if a testing effect would occur with an easier version of the weather prediction task.
Generally, Experiments 1 showed a benefit of testing, but no testing effect. Experiment 2 showed evidence of a testing effect with a task that require declarative memory. Experiment 3 did not show evidence for a testing effect in declarative memory due to overall low performance in the declarative condition. Showing a testing effect for the declarative conditions was important as a manipulation check and allowed for comparison between the nondeclarative and declarative conditions. If no testing effect was found in the declarative condition, then the results from the nondeclarative condition would be inconclusive. But because there was a testing effect in the declarative condition in Experiment 2, it is possible to investigate the effect of testing in the nondeclarative conditions as well. There was no testing effect found for the declarative conditions for Experiment 1 or 3.

**Testing effect and Nondeclarative Memory**

The focus of this dissertation is to investigate the effect that testing has on nondeclarative memory, specifically if testing decreases forgetting for nondeclarative memory. Before the results of the three experiments in this dissertation are discussed, it is important to highlight again the key differences between declarative and nondeclarative memory. Declarative memory is the acquisition, retention, and retrieval of information that can be intentionally recollected (Cogen & Squire, 1980). Declarative memory is associated with the hippocampus and medial temporal lobe structures. Nondeclarative memory includes the acquisition, retention, and retrieval of knowledge that is expressed by changes in performance and is not as easily consciously recollected (Squire & Dede, 2015). Depending on the type of nondeclarative memory, it is associated with the dorsal striatum, the neocortex, and the cerebellum (Packard, Hirsh, & White, 1989). Declarative and nondeclarative memory are largely distinct in both their cognitive functioning and the neural systems recruited.
Experiment 1 of the dissertation is the first research that investigates the testing effect for a task thought to recruit nondeclarative memory, the word fragment completion task. Comparable to the declarative condition, there was no traditional testing effect. Instead, there was an overall benefit of testing in that priming was greater for the tested words at the immediate and delay time points. These results do not suggest a testing effect for nondeclarative memory, but there was also not a testing effect for the declarative memory condition.

Like Experiment 1, Experiment 2 used the word fragment completion task with a declarative and nondeclarative condition. The two key differences between Experiment 1 and Experiment 2, are that in Experiment 2 feedback was not provided and participants were split into being “test aware” and “test unaware”. The results of experiment 2 show a testing effect in the nondeclarative condition. Not only was there less forgetting in the tested words, the priming did not decrease after 48 hours for only the tested words. While this lack of forgetting is not typically shown as a result of testing, this is still a benefit of testing and shows a testing effect. Participants who were aware that the words in the test phase came from the studied list of words were test aware used declarative memory to complete the word fragments and those that were test unaware used nondeclarative memory. The pattern of data for those that were test unaware mirrors the pattern of data for all participants, whereas the pattern for those that were test aware does not. This suggests that the interaction in the nondeclarative condition, that testing had a greater effect on priming for those in the delay condition, was due to those that were test unaware. This shows even greater evidence for a testing effect within nondeclarative memory, a completely novel finding.

Experiment 3 was designed to assess the effect that testing has on a different type of nondeclarative memory using the weather prediction task. The weather prediction task has been
used as a task that requires nondeclarative memory to perform optimally (Knowlton et al., 1994; 1996). The nondeclarative, feedback based, version of the weather prediction task has been shown to recruit the neural systems associated with procedural, nondeclarative memory (Poldrack et al., 2001). The results of Experiment 3 showed no difference in long term memory for the weather/pattern associations between the participants that learned by studying and those that learned by testing in the nondeclarative condition. The impact of testing was also evaluated for those that used a nondeclarative strategy to complete the task. Again, there was no difference for those that learned by studying or testing. Performance did not drop between the first day and 48 hours later. Because of this, it is inconclusive as to the effect that testing has on procedural nondeclarative memory.

**Types of Nondeclarative Memory**

The results from the three experiments in this dissertation did not show the same effect of testing on nondeclarative memory. This is likely since different types of nondeclarative memory are recruited for Experiment 1 and 2, and Experiment 3. Experiment 1 and 2 relied upon priming, which refers to the process in which experience increases the accessibility of information (Tulving and Schacter, 1990). Participants in Experiment 1 and 2 rated a series of words, and this experience leads to priming in the form of a higher likelihood that the participants would complete the word fragments for words they had seen as compared to novel word fragments. Priming is associated with decreases in activation in the bilateral extrastriate cortex, left fusiform gyrus, and bilateral inferior frontal gyrus during encoding (Schott et al., 2005), and decreased activation in the extrastriate visual cortex during retrieval.

Experiment 3 relied upon a different type of nondeclarative memory, procedural memory. Procedural memory relies on repeated experiences and enables the gradual learning of skills,
including navigation, and probabilistic categorization (Lum, Conti-Ramsden, Page, & Ullman, 2012). Procedural memory is associated with areas such as the cerebellum and the dorsal striatal brain systems (Packard, Hirsh, & White, 1989). In Experiment 3, participants in the testing condition learned with trial-and-error to associate different weather patterns with geometric cues.

The priming and procedural memory systems differ not only in the neural systems involved, but they also differ in how they interact with memory. Importantly, they differ in the stability of memory. In Experiments 1 and 2, there was a difference in priming from the immediate condition to the delay condition 48 hours later. In a similar task, the word stem completion task, priming decreased significantly after as short of a time as 30 minutes (Graf & Mandler, 1984; Graf et al., 1984; Endel Tulving, Schacter, & Stark, 1982). Some research, using the word fragment completion task, has shown priming after 7 days (McBride & Dosher, 1997). This research suggests that priming experiences decay in that priming decreases over time.

Procedural memory differs from priming in that procedural memory lasts for greater extended periods of time. Research on forgetting with procedural memory has used the serial reaction time task (Romano, Howard, & Howard, 2010). The serial reaction time task is a task that involves participants learning patterns by pressing one of four keys on a keyboard when a specific item appears on the screen (for reviews, see Forkstam & Petersson, 2005; Robertson, 2007). Learning is assessed over time be comparing the speed of learned patterns as compared to random patterns, despite participants not having any declarative knowledge of the patterns themselves. Retention of these patterns was shown at one-year post training, indicating that even with no opportunity to practice, procedural memory is exceptionally long lasting. The results from Experiment 3 are consistent with the work regarding the serial reaction time task.
The Effect of Feedback

One of the primary differences between Experiment 1 and 2 is that in Experiment 1 participants were given feedback during the test phase. While Experiment 1 and 2 differ in that in Experiment 2 participants were recruited from Amazon’s Mechanical Turk, the experiments were similar enough to roughly compare and investigate the effect that feedback has on the testing effect. First, it should be noted that priming was overall much higher in Experiment 1 than in Experiment 2. This is most apparent for the immediate condition (Experiment 1: Study Immediate M = 3.212, SD = 2.619, Test Immediate M = 11.636, SD = 2.247 and Experiment 2: Study Immediate M = 1.938, SD = 3.212, Test Immediate M = 4.479, SD = 3.620). This is likely due to the fact that feedback inherently increases the exposure of the tested words specifically. With feedback, if a participant incorrectly completes a word fragment or is unable to complete the word fragment, they are shown the correct work. Without feedback, they would not be shown the word if they fail to complete it. This extra exposure to the words with feedback likely explains the increase in priming for the words for Experiment 1.

Despite the difference in performance between Experiment 1 and Experiment 2, the overall trends within each experiment can be compared to understand how feedback impacted long term memory. The first thing that should be noted is that overall, there was greater priming for the tested words in Experiment 1 as compared to Experiment 2. This is likely because feedback gave extra exposure of the tested words specifically. In Experiment 1, participants saw the tested words once during the study phase, and once again during the test phase regardless if they correctly completed the word fragments or if they did not. In Experiment 2, participants only saw the tested words during the testing phase if they correctly completed the word fragment. If they did not complete the word fragment completely, they would not see the word a
second time. This extra exposure could cause greater memory of the words later in both the immediate and delay conditions. Therefore, one effect of feedback is that it increased the memory for the tested words overall.

The other question of interest is if feedback has any effect on the testing effect itself. The testing effect has been shown both in the presence of feedback (Kang et al., 2007) and in the absence of feedback (example: Roediger & Karpicke, 2006a). In Experiment 1, where feedback was provided, no testing effect was shown, only an overall benefit of testing. In Experiment 2, where feedback was not provided, a testing effect was shown. The presence of feedback was a primary difference between Experiment 1 and 2. Additionally, there was a greater benefit to priming for tested words in the immediate condition for Experiment 1 for the nondeclarative condition, and a greater benefit for tested words in the delay condition in Experiment 2. It is possible that the addition of feedback in experiment 1 caused greater priming for the tested words that was not sustained over a 2-day period. The feedback may have only provided short term benefits to priming. The results here are not enough to strongly suggest that feedback was the reason for no testing effect in Experiment 1, especially considering that the testing effect has been shown with feedback. But it is worth considering for future research.

Theories of the Testing Effect

The present dissertation is the first literature showing that testing results in less forgetting relative to studying for nondeclarative memory with perceptual priming specifically. The results discussed here have multiple important implications for existing testing effect theory. The first implication is on the testing effect itself. Previously, there had been no evidence of a testing effect for tasks that primarily require nondeclarative memory. This is important for considering the theories describing how the testing effect works, as detailed in Table 1. Earlier, theories of
the testing effect were detailed and used to make a prediction of the testing effect with nondeclarative memory. In the following section, these theories of the testing effect will be further discussed in how they pertain to the testing effect shown in Experiment 2 with perceptual priming.

Transfer appropriate processing, state-dependent learning between how information is encoded and how it retrieved, likely play role in the testing effect found for nondeclarative memory in Experiment 2 (Lockhart & Craik, 1972). Transfer appropriate processing has been thought to potentially cause a testing effect with declarative memory because of a similarity in the mental processes for a practice test and a final evaluative test. This similarity causes tested items to be remembered at a greater rate than studied items. A similar process would explain the testing effect found with perceptual priming in Experiment 2. Perceptual priming is sensitive to a match between the physical characteristics of what is primed and when it is retrieved after a delay (Blum & Yonelinas, 2001; Kinjo & Snodgrass, 2000; Roediger & McDermott, 1993; Rueckl & Mathew, 1999). This importance for a match is known as stimulus attribute sensitivity. In Experiment 2, there was an exact match for the tested words and the final evaluation phase that did not exist for the studied words. This exact match likely caused not only the testing effect found in Experiment 2, but also the overall benefit of testing found in Experiment 1 and 2.

Testing is generally more difficult than studying. This increased difficulty has been suggested to enhance the memory strength of tested items by causing deeper processing for the tested items (Landauer & Bjork, 1978). Deeper processing then leads to increased long-term memory of what is learned. Perceptual priming been shown to not be affected by levels of processing effects (Bowers & Schacter, 1990; Roediger et al., 1992). Therefore, even if testing
with a task that requires priming such as the word fragment completion task is more difficult than studying, this theory does not inform the testing effect found in Experiment 2.

The dual systems theory posits that testing creates a separate testing memory from a study memory created by studying (Rickard & Pan, 2018). During a final evaluative test, both the study memory and a testing memory can aid in retrieval. Items that are tested are benefited by having an original study memory that is formed during initial learning, and a separate testing memory that is created during a practice test. This theory makes no claim as to the actual mechanisms behind the creation of the study and test memory or the brain systems involved. It is entirely possible that a separate test and study memory would be created for the primed words found in Experiment 1 and Experiment 2. During the final evaluation phase, participants may have been primed to complete the tested words more than the studied words because of an additional priming test memory.

Another theory that has been posited to explain the mechanisms for the testing effect is the distribution-based bifurcation model (Halamish & Bjork, 2011; Kornell et al., 2011). The distribution-based bifurcation model assumes that testing increases the memory strength of tested items to a greater degree than studying increases the memory strength for studied items. The memory strength for these items (the studied and tested) are forgotten at similar rates, but because the tested items have greater memory strength they are more easily retrieved after a delay. A disproportionate increase in the memory strength for tested words, as compared to the studied words, would explain the benefit of testing for priming found in Experiment 1 and 2.

The Elaborative Retrieval hypothesis suggests that items semantically related to a target are recalled during testing (Carpenter, 2009). The greater the semantic distance between a cue
and a target leads to increased elaboration, which then leads to a greater testing effect. For example, consider the word pair “wheel-window”. The “wheel-window” word pair are connected by wheel, to car, to window. During the testing of “wheel-window”, the word “car” is also activated and during a final evaluation phase the words “car” may be an aid for the cued-recall of “window” when “wheel” is presented. These semantic elaborations are created to a greater extent during testing as compared to studying, and therefore leads to greater memory of tested items during a final evaluation phase. The testing effect found in the nondeclarative condition in Experiment two specifically involved perceptual priming. Perceptual priming involves the learning of the physical characteristics of an item, not the semantic information (Blum & Yonelinas, 2001; Kinjo & Snodgrass, 2000; Roediger & McDermott, 1993; Rueckl & Mathew, 1999). Since the Elaborative Retrieval hypothesis is based on semantic elaboration, it does not inform the testing effect found in this dissertation with perceptual priming.

Another more recent theory detailing a mechanism of the testing effect is the Episodic Context theory (Karpicke et al., 2014). This theory essentially states that testing involves attempting to reinstate a prior learning context. If retrieval is successful, the representation of the context, including temporal and semantic information, is updated to include features of retrieved contexts. This theory has been argued to account for the testing effect, the finding that testing impacts recall to a greater degree than for than recognition, the finding that spaced retrieval is better for learning, and that weaker cues produce a greater testing effect than strong cues. The idea behind this theory, is that the more contexts an item is retrieved in, the more features the item is associated with. This leads to increased contextual cues that can aid in final retrieval during a final test. When considering this theory as an explanation for the testing effect found in Experiment 2, or even the overall benefit of testing found in Experiment 1, it is important to
consider if different contexts would aid in perceptual priming. First of all, perceptual priming is increased with repeated exposure to a given stimulus (Tulving & Schacter, 1990), but the amount of exposure should be equal for words studied and words correctly retrieved during Experiment 1 and 2. If perceptual priming increases with varied exposure (the tested words) more than exposure of the same item (the studied words), then this theory may explain some of the testing effect found in Experiment 2. Perceptual priming relies on a close physical match between a learned stimulus and a response, therefore a variety in learned contexts would not aid in long term priming and this theory does not explain the testing effect found in Experiment 2.

To summarize, there was a testing effect shown in the nondeclarative memory condition in Experiment 2. This means that testing increased retention for priming, something that has never been shown prior to this dissertation. The exact mechanisms that have been used to explain a testing effect have been considered for if they explain a testing effect for nondeclarative memory, specifically priming since a testing effect was found with perceptual priming. Some of these theories make claims that testing causes deeper processing or greater elaboration, which would not affect the priming of the words in Experiment 2. Others focus more on a match in the mental processes between practice testing and a final evaluative phase, which would explain a testing effect for priming. It is clear that not all of the current testing effect theories extend to other types of memory that are not affected by deeper processing and semantic elaboration. In the next section, specific mechanisms of the testing effect are considered for declarative and nondeclarative memory.
Cognitive Mechanisms of the Testing Effect

The theories presented previously detail cognitive theories as to how testing improves long term retention and decreases forgetting. Cognitive explanations for the testing effect generally resolve around a few key ideas. One, memory representations change as a result of retrieval during testing due to elaboration of relevant information. Two, there is an inherent match between a practice test and a final evaluative test. The similarity in mental processes between the practice and final test results in greater recall of information during the final test. Third, retrieval during testing provides an additional context that can be used during retrieval during a final evaluative test. Finally, testing is more difficult and more effortful that results in deeper processing of the information. These cognitive explanations have been used to explain how testing improves long term retention by decreasing forgetting for things like memorized word pairs and may not apply to the testing effect found in Experiment 2.

It is possible that testing decreased forgetting for the primed words differently in Experiment 2 for the declarative condition as compared to the nondeclarative condition. In the declarative condition, forgetting may have been decreased for the tested words due to the theories listed above. In the nondeclarative condition, forgetting was likely decreased for the tested words for two reasons. The first is due to the match between the test in the training phase and the final evaluation phase that does not exist for the studied words. Perceptual priming heavily relies on a physical match between the item learned and what is retrieved (Blum & Yonelinas, 2001; Kinjo & Snodgrass, 2000; Roediger & McDermott, 1993; Rueckl & Mathew, 1999). Because of this, having a physical match between the tested words and the final evaluation phase would result in greater priming immediately and two days later, which was found in Experiment 1 and 2. Another cognitive mechanism at play is an increase in attention as
a result of testing. Perceptual priming requires some level of attention in order for priming to take place. During the study phase, participants rated the to be learned words in order to encourage that participants attend to the words. It is possible that there was greater attention during the testing in the training phase than the rating of the words in the study phase and this increased attention may have resulted in more priming. The cognitive theories described are not usually constrained by the way that the human brain actually works and overall have an abstract level of description.

**Neural Mechanisms of the Testing Effect**

In this section the neural mechanisms of the detailed in how they pertain to declarative memory and if they would explain the testing effect found for perceptual priming in Experiment 2. To date, there have been relatively few studies that have investigated the neural correlates of the testing effect. Four studies that have directly compared brain activity during practice testing and restudying showed lower activity in semantic storage areas (such as the left temporo-parietal areas) during a practice test as compared to studying (Rosner, Elman, & Shimamura, 2013; Van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013; Vannest et al., 2012; Wing, Marsh, & Cabeza, 2013). Differences in brain activation during the studying of items that are correctly retrieved later compared to items not correctly remembered is evidence of successful encoding. During testing, and not restudying, increased activity was found in these areas related to semantic memory storage such as inferior temporal gyrus and the middle temporal gyrus. These results suggest activity in semantic memory storage areas are different for restudying and testing and important for the testing effect. Overall, successful testing strengthens the neural representation of information in temporo-parietal areas (areas important for semantic memory)
whereas restudying evokes semantic information that is less relevant for learning (Van den Broek et al., 2016).

Previous research implicates neural areas important for semantic memory as important for the testing effect. A question to be considered is if activity in these areas would explain the testing effect within the nondeclarative memory condition in Experiment 2, or even the overall benefit of testing found in Experiment 1. Experiment 1 and 2 involved the word fragment completion task, a task thought to require perceptual priming (Roediger et al., 1992). Perceptual priming is associated with decreased activity in sensory cortices, for example the extrastriata cortex (cite squire 1992). This decreased activity is thought to be a result of fine tuning of neurons to the physical characteristics of what is primed. Perceptual priming occurs independently of areas associated with declarative memory as evident by individuals with impaired declarative memory performing normally on tasks that require priming (Graf, Squire, & Mandler, 1984). Therefore, the neural mechanisms important for the testing effect with declarative memory are likely largely different for perceptual priming.

Determining the exact neural mechanisms for a testing effect with nondeclarative memory is outside of the scope of this dissertation. Despite this, it is reasonable to assume that any neural mechanism of a testing effect with perceptual priming must lie within sensory areas. Testing likely results in greater decreased activity in the occipital lobe as compared to studying since decreased activity is associated with priming and testing led to increased priming for tested words in Experiment 1 and 2. There are a couple potential explanations for this, one being that participants may not have attended to the words during the study phase sufficiently to cause a fine tuning of neurons which would then not result in decreased activity in the occipital lobe and less priming. Another explanation is that testing causes a greater fine tuning of neurons in
sensory areas as compared to studying. This could be similar in nature to the testing effect mechanisms that cause a greater memory strength specifically for tested memory as compared to study memory. The goal of this dissertation was not to determine the exact neural mechanism behind a testing effect for nondeclarative memory, it was to determine if a testing effect exists for nondeclarative memory. Since a testing effect was found for one type of nondeclarative memory, perceptual priming, it is logical to assume that testing affects the neural representation of the primed words. Further research should be done to assess the neural mechanisms for how testing increases perceptual priming.

**Memory Systems**

Nondeclarative and declarative memory are not completely separate entities. That is, a task that is a “nondeclarative memory task” may also recruit declarative memory and a “declarative memory task” may recruit nondeclarative memory. Take the word stem completion task as an example. The word stem completion task is believed to involve priming, a form of nondeclarative memory. Like Experiment 1 and Experiment 2 in this dissertation, participants rate a series of words and then are shown incomplete words. The idea is that when participants are shown the incomplete words, they unconsciously are more likely to complete the incomplete words with words they had seen previously. There are a couple of issues with this assumption.

One issue with the assumption that participants would only use nondeclarative memory, is that some participants may realize that some of the word fragments came from the words they had seen earlier. These participants are said to be “test aware” and those that are test aware use declarative memory to complete the task (Bowers & Schacter, 1990). Obviously, this is an issue if the goal is to measure nondeclarative memory. The primary goal of Experiment 2 was to address some of these issues from Experiment 1. To closely mimic previous testing effect
studies, participants were given corrective feedback when completing the word fragments in Experiment 1. It is possible that this feedback could have caused participants to be more likely to consciously use the words that were rated in the study phase to complete the fragments. Therefore, the participants would be relying on declarative memory, even in the nondeclarative condition. For this reason, in Experiment 2 there was no feedback throughout the experiment. Also, in Experiment 2, participants were asked at the end of the experiment how they had completed the word fragments. Those that wrote that they used their memory of the previously seen words were considered “test aware” and a testing effect was seen even for those that were “test unaware”.

An area of research within psychology is whether human learning and memory is mediated by a single system or by multiple systems. This dissertation has argued for distinct memory systems, specifically a declarative memory system and a nondeclarative memory system. In the area of memory research, it is largely accepted that there are multiple memory systems (Cohen & Squire, 1980; Memory & Endel Tulving, 1987; Packard et al., 1989; Squire et al., 1992). Experiment 1 and 2 used priming tasks to measure the testing effect. Priming has been used as a primary example of multiple memory systems by showing that despite extreme deficiencies with declarative memory, individuals with amnesia can still show normal levels of priming (Hamann & Squire, 1997). Despite this evidence, a single-system model of priming has been formalized that is claimed to explain intact priming and impaired declarative memory in individuals with amnesia (Berry, Kessels, Wester, & Shanks, 2014; Berry, Shanks, & Henson, 2008a, 2008b; Berry, Shanks, Li, Rains, & Henson, 2010; Shanks & Berry, 2012).

If there was a single system for memory, it could be expected that testing would have a similar impact on declarative and nondeclarative memory. In Experiments 1, 2, and 3, there was
no significant difference in how testing impacted memory between the declarative and nondeclarative memory conditions. This does not necessarily give evidence to a single-system theory of memory. A multiple memory system approach could yield comparable results between declarative and nondeclarative memory. The data in the present dissertation are equivocal with regards to the memory systems debate.

**Limitations**

One major limitation for Experiment 1 and Experiment 2, is that declarative memory may be used even when the participant is not aware that they are using the words from the study phase during the final evaluation phase. This can occur even without effortful or deliberate recollection (Gardiner, Java, & Richardson-Klavehn, 1996; Schacter & Tulving, 1987; Richardson-Klavehn, Lee, Joubran, & Bjork, 1994; Schacter, Bowers, & Booker, 1989). This contamination of explicit memory was shown in the first neuroimaging study of priming (Squire et al., 1992). In this experiment, participants studied a list of words and completed word stem completions during three PET scans. In one scan, the stems were from the list studied (priming). In another, the stems came from only novel words (baseline). In the third scan, participants were explicitly told to use the words they had seen before (explicit recall/declarative memory). Priming was associated with decreased activity in the right extrastriate occipital cortex. Interestingly, there was also activity in the hippocampus during priming. Considering research showing that priming is possible with a damaged hippocampus, it is likely that this activation reflects the use of declarative memory, intentional or not. Therefore, it is possible that participants that were even considered test unaware used nonconscious declarative memory to complete some word fragments.
Experiment 3 provided its own set of limitations. One such limitation was mentioned earlier, and that is that participants did not use declarative strategies to a great enough degree to measure a testing effect with declarative memory. In Experiment 3, less than 20% of participants used a declarative strategy that was not simply guessing. Previous research that used similar models to assess the participants strategy found that 70% percent of the participants used a declarative strategy after the fourth training block (Gluck et al., 2002). The significant difference here is interesting because Experiment 3 of this dissertation and the study by Gluck and colleagues, 2002, shared similar stimuli and methodology. The only major difference is that the study by Gluck and colleagues modeled the participants decision strategies during the fourth training block whereas the participants in Experiment 3 had their decision strategies modeled during the fifth block (the final test block). The reason for this is because there is no data for the participants in the study condition to use before the test block. A better comparison would be with the decision models of the participants in the test condition during the fourth training block, which had 61% of participants using an optimal strategy. While it isn’t clear why so few participants used a declarative strategy, there were not enough participants to assess the effect of testing for participants with a declarative strategy.

Another limitation of the weather prediction task comes from using the decision strategy models to classify participants into using strategies that are either declarative or nondeclarative in nature. It is possible that a participant uses multiple strategies during the task. Modeling their best fitting model only gives the strategy they had most likely used. Also, it is assumed that participants who used a multi-cue model relied on nondeclarative memory or those that used something like a single-cue or one-cue relied on declarative memory. The reason for this assumption is because use of declarative strategies is associated with hippocampal activity and
dorsal striatal activation is associated with the use of nondeclarative strategies (Schwabe & Wolf, 2012). It is possible for someone to memorize all 14 patterns and the optimal strategy and use declarative memory to solve the task and its possible that someone learns incrementally with nondeclarative memory to use one of the less optimal strategies.

**Future Directions**

This dissertation features novel research investigating the effect that testing has on nondeclarative memory. More research is necessary and this section details suggestions for future directions. First, future directions will be considered regarding the use of tasks like the word fragment competition to measure the effect that testing has on priming. Second, future directions will be considered for the use of procedural tasks like the weather prediction task to study the testing effect. Finally, future directions will be considered for other types of nondeclarative memory as well.

Experiment 1 and 2 were the first experiments to use the word fragment completion task to study the testing effect with priming. Future research should focus on ensuring that participants are correctly classified as using nondeclarative or declarative memory. There are multiple ways to do this, one of which involves the use of neuroimaging. Previous research has used different neuroimaging techniques (example, PET and FMRI) to measure the amount of hippocampal activation during word fragment and word stem completions (Schott et al., 2005). Hippocampal activation is associated with the use of declarative memory and word fragment completions with increased hippocampal activity are associated with the use of declarative memory, conscious or not. If testing increases the memory for fragments that are completed without additional hippocampal activity, that would be greater evidence for a testing effect with nondeclarative memory.
Additionally, procedures can be used to help determine if a specific word fragment was recalled with primarily declarative or nondeclarative memory. In one example of such research, participants indicated whether or not they remembered that an item during the completion phase had appeared earlier in the experiment (Gardiner et al., 1996; Gardiner, Ramponi, & Richardson-Klavehn, 1998). The goal of this was to determine on an item by item basis if declarative memory was used. Experiment 2 only attempted to determine overall test awareness, instead of measuring on an item by item basis. Also, different procedures may limit the number of participants that become test aware. This could be something like increasing the amount of novel words during the final evaluation phase, so participants are not seeing fragments for words they had seen earlier in as high of a concentration. The goal of these proposed future directions is to both limit the number of participants using declarative memory in the nondeclarative condition and to better detect the influence of declarative memory in the nondeclarative condition.

There are two primary limitations that future research should address with using the weather prediction task to study the testing effect. First, the results of Experiment 3 showed that an exceedingly small proportion of participants were using a declarative strategy during the final evaluation phase and those that did showed performance that is too low to detect a testing effect. While the probabilities in Experiment 3 were already adjusted from the original weather prediction task, it is possible that adjusting them to make the task easier would encourage more participants to adapt a declarative strategy. The probabilistic nature of the task is one of primary reasons that it can require procedural memory, so changing the probabilities to be easier to predict the weather pattern may lead to more declarative strategy use (Gluck et al., 2002; Knowlton et al., 1996, 1994). Also, performance did not drop from the immediate to the delay time points for participants that used a nondeclarative strategy regardless of training type. While
this may be the nature of the that being procedural in nature, it is possible that increasing the delay period could cause a decrease in performance.

This dissertation used two different tasks, the word fragment completion task and the weather prediction task, tasks of priming and procedural memory respectively. Future research should expand on this research by using other nondeclarative memory tasks. This could include other types of priming tasks, like the word stem completion task or the lexical decision task. It could also include other types of procedural memory tasks, such as the serial response time task. Also, future research could investigate the effect on other forms of nondeclarative memory, such as associative learning like classical conditioning and non-associative learning. Future research should also investigate the real-world implications of a testing effect with nondeclarative memory. This could look like having students in a classroom learn tasks thought to require nondeclarative memory, and then either study or be tested on these tasks and measure their performance in actual classroom exams. This dissertation is just the beginning of an entire area of research that had previously not been studied and there is much more research to be done in this area.
General Experiment Conclusions

The findings in this dissertation contribute to both the testing effect literature and the memory systems literature. Key questions were asked in the introduction of this dissertation. 1) Does studying cause immediate enhancement to memory? The experiments in the dissertation do not support an immediate benefit of studying. 2) Does testing reduce the amount of forgetting relative to studying? The results from Experiment 2 suggest that testing can decrease forgetting relative to studying. 3) Does the type of learning strategy reduce the amount of forgetting relative to studying? The results from Experiment 3 suggest that depending on the task, only specific learning strategies result in learning.

The most prominent finding was a testing effect for nondeclarative memory in Experiment 2. In Experiment 2, there was not only a benefit of testing in that priming for the tested words were higher, but there was no decrease in priming after two days when for the studied items performance decreased. These results provide support for a testing effect for nondeclarative memory, specifically priming. Experiment 3 suggests that procedural memory may not benefit from testing during the period of only two days. It is likely that procedural memory is long lasting regardless of if it was learned by studying or by testing. Further research needs to consider either a longer time period or a task that results in a decrease in procedural memory over two days. Additionally, it may be better to conceptualize how testing may benefit long term procedural memory differently than simple decreased forgetting. Testing may increase the speed of learning or may increase the transfer of what is learned.

The results presented in this dissertation shows the importance of studying how testing impacts nondeclarative memory. The first reason this research is important is due to how prevalent testing is in educational settings. Testing has been accepted as a norm for classrooms
across the world, as it not only shows the competency of students, but also increases the memory of the learned information. Previously, it was a well-studied phenomenon that testing improves declarative memory, memory for things like memorized facts. Now, these results show that testing can be a useful tool for things that involve nondeclarative memory. For example, consider a student learning a new language. Nondeclarative memory has been implicated in the use and the learning of language (For review see Ettlinger, Margulis, & Wong, 2011). This dissertation suggests a greater benefit if the student learns the language by testing themself on the new language, instead of simply studying the words and grammar of that language.

Previous research on the mechanisms of the testing effect have solely used tests that require primarily declarative memory, such as the memorization of word pairs or information in prose passages. Because of this, mechanisms have been described that may only explain a testing effect with very explicitly learned material. This dissertation shows that testing impacts nondeclarative memory differently than declarative memory. Also, testing impacts distinct types of nondeclarative memory uniquely, as evidenced by the differences in the results of Experiment 2 and Experiment 3. As a result, when investigating the mechanisms of the testing effect it is important to not only consider if the primary memory type recruited is declarative or nondeclarative, but also what type of nondeclarative memory is recruited.

This results from this dissertation show that it is important to consider the differences between distinct types of nondeclarative memory. Often, memory is thought of as either declarative or nondeclarative memory. But this dissertation suggests that memory should be thought of at least as declarative, procedural, priming, etc. Also, learning should not be viewed as requiring either declarative or nondeclarative since both declarative and nondeclarative memory are required during learning. The Experiments used in this dissertation were chosen
because they are thought to primarily require nondeclarative memory, but even then, declarative memory may have been involved. Overall, this dissertation presents completely novel findings showing that testing can increase nondeclarative memory. This research should be used to inform education practices in classroom settings and is further evidence for benefits of testing.
References


Sutherland, R. J., & Rudy, J. W. (1988). Place learning in the Morris place navigation task is impaired by damage to the hippocampal formation even if the temporal demands are reduced. *Psychobiology, 16*(2), 157–163. https://doi.org/10.3758/BF03333120


Appendix

A. The word list used in Experiment 1, adapted from (Roediger et al., 1992).

<table>
<thead>
<tr>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>admiral</td>
</tr>
<tr>
<td>adultery</td>
</tr>
<tr>
<td>alias</td>
</tr>
<tr>
<td>alligator</td>
</tr>
<tr>
<td>anecdote</td>
</tr>
<tr>
<td>armadillo</td>
</tr>
<tr>
<td>ashore</td>
</tr>
<tr>
<td>bacteria</td>
</tr>
<tr>
<td>barnacle</td>
</tr>
<tr>
<td>beggar</td>
</tr>
<tr>
<td>behold</td>
</tr>
<tr>
<td>betray</td>
</tr>
<tr>
<td>blessing</td>
</tr>
<tr>
<td>butcher</td>
</tr>
<tr>
<td>cannon</td>
</tr>
<tr>
<td>caravan</td>
</tr>
<tr>
<td>chimney</td>
</tr>
<tr>
<td>crucifix</td>
</tr>
<tr>
<td>damsel</td>
</tr>
<tr>
<td>deceive</td>
</tr>
<tr>
<td>diamond</td>
</tr>
<tr>
<td>digest</td>
</tr>
<tr>
<td>dragon</td>
</tr>
<tr>
<td>epitaph</td>
</tr>
<tr>
<td>evergreen</td>
</tr>
<tr>
<td>feather</td>
</tr>
<tr>
<td>flourish</td>
</tr>
<tr>
<td>franchise</td>
</tr>
<tr>
<td>gauntlet</td>
</tr>
<tr>
<td>glitter</td>
</tr>
<tr>
<td>goddess</td>
</tr>
<tr>
<td>helmet</td>
</tr>
<tr>
<td>heredity</td>
</tr>
<tr>
<td>impetus</td>
</tr>
<tr>
<td>incline</td>
</tr>
<tr>
<td>inspire</td>
</tr>
<tr>
<td>maiden</td>
</tr>
</tbody>
</table>
malaria
mermaid
mischief
molecule
monarch
nutrient
oblivion
panorama
perfume
prisoner
projectile
publish
regency
sailor
scarlet
slumber
spinach
sprinkle
stanza
surround
thicket
torment
traitor
typhoon
twinkle
vanish
B. Strategy Analysis

Four different basic classes of strategy that subjects may use were investigated. A multi-cue strategy, in which the subject attends to all four cues. A guessing strategy, in which the subject guesses on each trial. One-cue strategies, in which a subject attends to only a single cue. Lastly, singleton strategies, in which the subject learns how single cues predict the outcome and guess whenever multiple cues are present.

Ideal data was created for each strategy. This data was created by the pattern of responses expected for each of the 200 trials as if a subject were to perfectly and reliably follow that strategy. These ideal data provide models of performance to be compared against the actual subject responses. Each subject’s data were fit to each of the created model data by taking the squared difference between the number of sun responses the subject produced and the number of sun responses predicted by the model, summed across all patterns. This was done for each 50 trial block during training and for the 200 trial test phase. This score was then normalized by dividing between the sum of squares of total presentations of each pattern.

\[
Score\ for\ Model\ M = \frac{\sum_p (Number\ of\ sun\ expected_{p,M} - Number\ of\ sun\ actual_p)^2}{\sum_p (Number\ of\ presentations_p)^2}
\]

The resulting score was a number from 0 to 1 for each model. A 0 indicates a perfect fit between a specific model strategy and a participant’s response. The lowest score for each model for each participant was determined to be the strategy used by that participant.
Biography of the Author

David Smith was born in Grand Rapids, Michigan on March 25, 1993. He was raised in Byron Center, Michigan and graduated from Byron Center High School in 2011. He attended Michigan State University and graduated in 2014 with a bachelor’s degree in Psychology. After receiving his degree, David will be taking part in the Data Incubator data science fellowship and will be continuing to a career in data science. David is a candidate for the Doctor of Philosophy degree in Psychology from the University of Maine in December 2020.