Does testing enhance mediation in paired-associate learning?

Deana Vitrano

University at Albany, State University of New York, vitrano.deana@gmail.com

The University at Albany community has made this article openly available.

Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Cognitive Psychology Commons

Recommended Citation

https://scholarsarchive.library.albany.edu/legacy-etd/2603

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive.
Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
DOES TESTING ENHANCE MEDIATION IN PAIRED-ASSOCIATE LEARNING?

by

Deana Vitrano

A DISSERTATION
Submitted to the University at Albany, State University of New York
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

College of Arts and Sciences
Department of Psychology
May 2020
# TABLE OF CONTENTS

Acknowledgements ........................................................................................................ iii

List of Tables .................................................................................................................. v

List of Figures ............................................................................................................... vi

Abbreviations ............................................................................................................... vii

Abstract ...................................................................................................................... viii

Introduction ................................................................................................................ 1

  Mediator Effectiveness Hypothesis ........................................................................... 5

  Mediators Shift Hypothesis ...................................................................................... 12

The Current Study ....................................................................................................... 13

Experiment 1 .............................................................................................................. 16

Experiment 2 .............................................................................................................. 21

Experiments 3a and 3b .............................................................................................. 26

  Experiment 3a .......................................................................................................... 28

  Experiment 3b .......................................................................................................... 37

General Discussion .................................................................................................. 46

References ................................................................................................................. 54

Appendices ................................................................................................................ 64

Tables ......................................................................................................................... 66

Figures ....................................................................................................................... 72
ACKNOWLEDGEMENTS

My heartfelt thanks to my advisor, Dr. James Neely, for the tremendous time and effort he has put toward helping me with this project and for all the assistance he has given me throughout my graduate career. His constant support and guidance have been absolutely invaluable to me. I would also like to express my appreciation to Dr. Jeanette Altarriba for her insightful guidance, unwavering support, and for serving as a member of my dissertation committee. I am also grateful to Dr. Heather Sheridan for serving as a member of my dissertation committee and for providing me with thoughtful feedback. Likewise, a sincere thanks should go to Dr. Bruce Dudek who has spent quite a bit of time teaching me statistical analyses and helping me prepare for a productive career, and to Dr. Tram Neill and Dr. Laurie Feldman for their courses and helpful feedback that helped me get to this point in my career.

Moreover, I would like to extend my gratitude to my fellow members of the laboratory, both past and present, Dr. Kit W. Cho, Stephanie Crocco, and Carol Bolte. My success would not have been possible without the knowledge they shared with me, the advice they gave me, and the time they spent with me in the laboratory. I am also extremely grateful for my fellow colleagues, particularly Jennifer Martin and Abigail Kleinsmith, who have taught me so much throughout my years at the University at Albany. Additionally, I had the pleasure of working with many hard-working undergraduate research assistants, who helped me conduct all of the experiments in my dissertation.

I also gratefully acknowledge the assistance of Casey Kohler, Brian Gabriel, Dr. JoAnne Malatesta, and the Office of Undergraduate Education in helping me reach my goals. I am especially appreciative of their constant encouragement throughout this process.
Lastly, I extend a large thank you to all of my loving family and friends. I thank my brother, Matthew, for his moral support and astute editing, and my parents, who always unconditionally supported all of my dreams and taught me to aim high. My success has been possible because of all of you!
LIST OF TABLES

Table 1. Mean Session 1 correct recall percentages for the three review phases………………….66
Table 2. Mean Session 2 final test correct recall percentages of mediators and targets to their corresponding cues…………………………………………………………………………………67
Table 3. Mean Session 2 recall percentages of a target to its cue, conditionalized on recall of the mediator (M). …………………………………………………………………………………………………………68
Table 4. Mean percentages of times the “specified” mediator was reported……………………69
Table 5. Total percentages of trials with mediator shifts, broken down by whether the shift was from a “specified mediator” or from an “unspecified” mediator………………….70
Table 6. Total percentages of trials with mediator shifts, broken down by the test versus restudy groups, and whether or not the target was successfully recalled to its cue in the immediately preceding phase……………………………………………………………71
LIST OF FIGURES

Figure 1. Procedures for Experiments 1 and 2 .........................................................72

Figure 2. Procedures for Experiment 3 .................................................................73
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAS</td>
<td>Backward associative strength</td>
</tr>
<tr>
<td>FAS</td>
<td>Forward associative strength</td>
</tr>
<tr>
<td>LSA</td>
<td>Latent semantic analysis</td>
</tr>
<tr>
<td>MEH</td>
<td>Mediator Effectiveness Hypothesis</td>
</tr>
<tr>
<td>MSH</td>
<td>Mediator Shift Hypothesis</td>
</tr>
<tr>
<td>TE</td>
<td>Testing effect</td>
</tr>
<tr>
<td>TPL</td>
<td>Test-potentiated learning</td>
</tr>
</tbody>
</table>
Abstract

The testing effect (TE) is the robust finding that testing on previously studied material leads to better long-term retention as compared to restudying that material (Roediger & Karpicke, 2006b). Pyc and Rawson (2010) proposed the Mediator Effectiveness Hypothesis (MEH) as an explanation for the TE in paired-associate learning. The MEH states that review testing on cue-target word pairs strengthens semantic/associative mediators, which helps participants recall targets to their cues on a later test. Pyc and Rawson found support for the MEH with Swahili-English word pairs and explicit mediation instructions, using the most rigorous test of the MEH. Other researchers have claimed support for the MEH with semantically related English word pairs and spontaneous mediation conditions (e.g., Carpenter, 2011), but they did not use the most rigorous test of the MEH. Using the most rigorous test with spontaneous mediation conditions and unrelated English word pairs for which semantic mediation and phonological mediation strategies should have been especially beneficial, the results of the current Experiments 1 and 2 failed to support the MEH. When semantic/associative or phonological mediation was explicitly encouraged in Experiments 3a and 3b, respectively, the MEH was supported only for phonological mediation. This was also the case for Pyc and Rawson’s (2012) Mediator Shift Hypothesis in that for phonological mediation, the tested group was more likely to shift mediators than the restudy group and particularly so after target recall failures on the immediately preceding test.

Keywords: testing effect, paired-associate learning, mediator effectiveness hypothesis
Does Testing Enhance Mediation in Paired-Associate Learning?

The testing effect (TE) is the finding that being tested on previously studied material leads to better long-term retention as compared to restudying that material (Roediger & Karpicke, 2006b). Studies on the TE typically utilize three distinct phases: (1) an initial study phase in which participants learn a set of materials, (2) a review phase in which half of the participants are given a test on the initially learned materials (test group) and the other half of the participants restudy the initially learned materials (restudy group), and 3) a final test in which all participants are tested on the initially learned materials. On the final test, the test group typically outperforms the restudy group (see Rowland, 2014, for a review).

The TE is a robust phenomenon and has been replicated numerous times in the laboratory with different populations, including children (e.g., Lipowski, Py, Dunlosky, & Rawson, 2014), college students (e.g., Buchin & Mulligan, 2019; Roediger & Karpicke, 2006a), older adults (e.g., Meyer & Logan, 2013), and even individuals with traumatic brain injuries (e.g., Pastötter, Weber, & Baüml, 2013; Sumowski et al., 2010). Furthermore, the TE emerges with a range of materials, such as single word lists (e.g., Carpenter & DeLosh, 2006; Rowland & Delosh, 2014; Rowland, Littrell-Baez, Sensenig, & DeLosh, 2014; Zaromb & Roediger, 2010), text passages (e.g., Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Blunt & Karpicke, 2014; Karpicke & Blunt, 2011), foreign language vocabulary word pairs (e.g., Carrier & Pashler, 1992; Karpicke & Roediger, 2008; Py & Rawson, 2007; Py & Rawson, 2010; Toppino & Cohen, 2009; Wilkinson, Hall, & Hogan, 2019), English word pairs (e.g., Carpenter, 2009; Carpenter, Pashler, & Vul, 2006; Cho, Neely, Brennan, Vitrano, & Crocco, 2017), face-name pairs (e.g., Carpenter & Delosh, 2005) and even nonverbal materials (e.g., Carpenter & Pashler, 2007; Kang, 2010; Wheeler & Roediger, 1992).
The TE has also been demonstrated with different test formats, including cued-recall (e.g., Carpenter, 2009; Carpenter, 2011; Pyc & Rawson, 2010), free recall (e.g., Roediger & Karpicke, 2006a; Zaromb & Roediger, 2010), and multiple-choice (e.g., Odegard & Koen, 2007). Additionally, the review test format does not need to match the final test format for the TE to emerge (e.g., Carpenter & DeLosh, 2006; Odegard & Koen, 2007; Roediger & Marsh, 2005; Veltre, Cho, & Neely, 2014).

Despite the TE persisting across different populations, materials, and test formats, the retention interval between the review phase and final test can affect whether or not a TE occurs. For example, the TE seems to reliably emerge when this retention interval is long (e.g., 30 min: Carpenter, 2011; 2 days: Roediger & Karpicke, 2006a; 7 days: Pyc & Rawson, 2010; Roediger & Karpicke, 2006a; Toppino & Cohen, 2009). However, when the retention interval is shorter (5 min), TEs can be null or sometimes even reversed TEs, in which restudying produces a greater benefit for subsequent memory than does testing (e.g., Roediger & Karpicke, 2006a; Toppino & Cohen, 2009; Wheeler, Ewers, & Buonanno, 2003).

**Real-World Settings**

The TE is not restricted to laboratory settings, but replicates rather reliably in real-world settings. For example, McDaniel, Anderson, Derbish, and Morisette (2007) examined the TE in a Brain and Behavior college course by giving weekly quizzes on assigned classroom readings. The classroom readings were approximately 40 pages of material from the textbook associated with the course (e.g., “All preganglionic axons, whether sympathetic or parasympathetic, release acetylcholine as a neurotransmitter”). The students received a different quiz each week for three weeks: one quiz with multiple-choice questions, one quiz with short answer questions, and one “quiz” that had no questions, but instead was comprised of sentences for the students to re-read.
All students received the three different types, counterbalanced for order. After three weeks of the weekly quizzes, all students took a multiple-choice unit test on the information from the first three weeks, and then repeated the entire procedure (i.e., three weeks of weekly quizzes and a unit test). Following the second round, all students received a final cumulative multiple-choice test on information from the entire six weeks. On both the unit tests and the final test, McDaniel et al. found that the short answer review questions enhanced recall relative to review re-reading whereas multiple-choice review questions did not. The findings demonstrate that the TE can extend from the laboratory to a real-world college course.

Numerous other studies have replicated the TE in an undergraduate student setting (e.g., Batsell, Perry, Hanley, & Hostetter, 2017; Bego, Lyle, Ralston, & Hieb, 2017; Cadaret & Yates, 2018; Carpenter, Rahman, & Perkins, 2018). Other studies have found the TE in a classroom setting for medical and dental students (Jackson, Hannum, Koroluk, & Proffit, 2010; Logan, Thompson, & Marshak, 2011), pharmacy students (Hernick, 2015), high school students (Achord, 2015; Evans, 2013; McDermott, Agarwal, Antonio, Roediger, & McDaniel, 2014), middle school students (Carpenter, Pashler, & Cepeda, 2009; McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013; McDermott et al., 2014; Roediger, Agarwal, McDaniel, & McDermott, 2011), elementary school children (Aslan & Bäuml, 2015), and even preschool-aged children (Kliegl, Abel, & Bäuml, 2018). Clearly, the TE is relevant in educational settings of various ages, and can be used as a tool to enhance student success in the classroom. Therefore, the TE is an important effect to continue to research in the laboratory and to understand theoretically.
Elaborative Retrieval Hypothesis

Despite the robustness of the TE both in the laboratory and the classroom, the best explanation of the TE has not yet been established. Carpenter (2009; see also Carpenter & DeLosh, 2006) proposed the elaborative retrieval hypothesis to explain the TE in paired-associate learning (i.e., cue-target word pairs). Carpenter (2009) argues that similar to spreading activation theories of memory (e.g., Anderson, 1983), a cued-recall test on a cue-target pair could activate other related, relevant concepts, as the participant attempts to recall the target. For example, Carpenter (2009) suggests that if the cue-target word pair is “basket-bread,” trying to retrieve “bread” could activate related words to the pair, such as “eggs” and “flour,” as the participant searches for the target “bread.” If this is the case, participants would then create more pathways to activating the correct target in a later memory test. Carpenter argues that this would likely not happen with participants who simply restudy the word pairs, because if the target is already present, there would be no reason to search for the target and activate other related words.

Furthermore, Carpenter (2009) argues that if the elaborative retrieval hypothesis is correct, a larger TE should emerge when the cue-target pairs are weakly related as compared to strongly related. That is, if the cues and targets are only weakly related to each other, the participants would need to activate more related concepts in order to arrive at the target. In two experiments, Carpenter used both strongly-related (e.g., toast-bread) and weakly-related (e.g., basket-bread) word pairs to examine the elaborative retrieval hypothesis. Participants first learned the word pairs in a self-paced study session, then received a short distractor task, followed by a review phase in which they were given a cued-recall test on the word pairs without feedback and/or restudied the word pairs. (Experiment 1 was a within-participants design for the
review phase whereas Experiment 2 was a between-participants design.) Following the review phase, participants engaged in a five-minute distractor task and then completed the final cued-recall test on the word pairs. Carpenter found a significant TE overall, and furthermore, found that the TE was larger with the weakly-related word pairs. Specifically, being tested on the weakly-related word pairs led to better long-term retention as compared to testing on the strongly-related word pairs, which Carpenter attributed to the greater elaboration needed for successful retrieval of a weakly-related word pair. In contrast, there was no difference in retention when the word pairs were restudied, likely because elaboration is not needed while restudying the pairs. Carpenter concluded that elaborative retrieval may be one mechanism underlying the TE.

The Mediator Effectiveness Hypothesis

Although there has been support for the elaborative retrieval hypothesis (e.g., Carpenter, 2009; Carpenter & DeLosh, 2006; Rawson et al., 2015), Carpenter (2011) points out that none of the studies specified exactly what elaborative information is becoming activated during retrieval practice. A more specific explanation is necessary to understand the role of elaborative activation in TEs. Pyc and Rawson (2010) proposed the mediator effectiveness hypothesis (MEH) as a more specific explanation for the TE for paired-associates (i.e., cue-target word pairs). They proposed that being tested on the cue activates a mediator, and the mediator in turn helps with recalling the target. Therefore, according to the MEH, testing strengthens the links of a cue-mediator-target chain.

Pyc and Rawson (2010) used Swahili-English word pairs (e.g., mashua-boat) and instructed participants to generate their own mediators that are phonologically similar to the cue and semantically related to the target (e.g., “mast”). Specifically, participants first completed a
study phase in which they were presented with 24 cue-target word pairs, and were told to overtly report a mediator. They then either restudied each pair (restudy group) or were tested on their recall of each target to its cue with the correct pair displayed immediately afterward (test-restudy group). Participants were once again told to overtly report a mediator on each restudy or test trial, and they were given the opportunity to change their mediators for each pair. After participants restudied or were tested on all 24 cue-target word pairs, they repeated the same restudy/test cycle two more times. Each time, they were again to overtly report a mediator on each study, restudy or test trial and were given the opportunity to change their mediators for each pair across trials.

One week later, all participants received one of three cued-recall final tests in which they were presented with: (1) the cue and asked to recall the target, (2) the cue and their most recently generated mediator and asked to recall the target, or 3) the cue and asked to recall both their most recently generated mediator and the target. Pyc and Rawson (2010) found a standard TE for cue-target recall, and found that participants in the test-restudy group had higher overall recall of the mediators as compared to the restudy group, suggesting that testing strengthened the cue-mediator link or made the mediator generally more accessible. Additionally, recall for the test group was the same whether participants were prompted with only the cue or both the cue and their most recently generated mediator, suggesting that the test group was covertly recalling their mediators when prompted with the cue only. More important for the MEH, a TE occurred for cue-target recall when the mediator could be recalled, but not when it could not be recalled. This strongly supports the MEH, and the idea of testing strengthening the links of a cue-mediator-target chain.
Spontaneous Mediation

Pyc and Rawson (2010) supported the MEH when explicitly instructing participants to use mediation, but it is not clear if spontaneous mediation would occur in paired-associate learning. Carpenter (2011) conducted two experiments examining spontaneous mediation with semantically related English cue-target word pairs (e.g., mother-child), each with a primary associate (mean forward associative strength of .57) of the cue that could serve as a mediator (e.g., “father”) for target recall. In Carpenter’s Experiment 1, participants first studied the English word pairs, completed a 20-second distractor task, and then either restudied or were tested on the same pairs. They were not required to overtly report any mediators. Following another short distractor task (five minutes), they then received a recognition test on the individual cues and targets that had appeared in the previously studied word pairs (old items), the putative semantic mediators that were not presented in the prior study list, and unrelated lures (e.g., “bread” for mother-child) that were also not presented in the prior study list (both new items). Carpenter found that false alarms were overall higher for the mediator lures than for the unrelated lures and that false alarms were higher to the semantic mediators (but not to the unrelated lures) when the review phase involved a test rather than restudy. This shows that testing activated the semantic mediators more than restudying.

Carpenter’s (2011) Experiment 2 procedure was the same as that of Experiment 1 except that the final test was a cued-recall test instead of a recognition test and the retention interval between the review phase and final test was 30 minutes rather than 5 minutes. In the cued-recall final test, participants were prompted with either (a) a previously studied cue, (b) a non-studied word that was related to the cue and target but was not a primary associate of the cue and hence would have been unlikely to serve as a mediator (e.g., “birth” for mother-child), or the putative
semantic mediator (‘father’ for mother-child). There was a TE for target recall to the semantic mediator but not for target recall to the ‘nonmediator’ related cue. Carpenter concluded that these results supported the MEH.

**Replications of Carpenter (2011).** Since Carpenter’s (2011) studies, there have been several replications of her Experiment 2. Rawson et al. (2015) conducted two experiments that were nearly identical to Carpenter’s (2011) Experiment 2, and replicated Carpenter’s (2011) major findings. That is, in support of the MEH, Rawson et al. found a significantly larger TE for target recall when the participants were cued with the mediators (e.g., pig for pork-beef) as compared to related words (e.g., cattle).

Similarly, Coppens, Verkoeijen, Bouwmeester, and Rikers (2016) performed several studies that were aimed at replicating Carpenter’s (2011) Experiment 2, with the only major difference being that their participants were Amazon Mechanical Turk workers, instead of undergraduate students. The materials were word pairs with similar characteristics as the word pairs used in Carpenter (e.g., sea-river, mediator: ocean). Coppens et al.’s (2016) Experiments 1 and 2 failed to replicate Carpenter’s (2011) finding that a larger TE occurs for the target recall to the mediators as compared to target recall to the related words; however, Coppens et al.’s (2016) Experiment 3 replicated the finding. That is, in Experiment 3 (a replication of the previous experiments), there was a larger TE for mediator-target recall as compared to related word-target recall, although it was a much smaller effect than in Carpenter’s (2011) Experiment 2. Cho, Neely, Brennan, et al. (2017) found a similar result to Coppens et al.’s (2016) Experiment 3 using similar word pairs in their replication. Congruent with Carpenter’s (2011) findings, Cho, Neely, Brennan, et al. (2017) found a significant (but substantially smaller) TE for mediator-target recall, but a nonsignificant TE for related word-target recall.
**Backward-Chaining-Hypothesis.** Both Coppens et al. (2016) and Cho, Neely, Brennan, et al. (2017) independently assessed an alternative, backward-chaining account of why mediator-target recall showed a larger TE than related word-target recall in some of the above experiments. The word pairs in Carpenter (2011) had a strong mediator-cue backward association (e.g., for “mother-**father**-child”, **father**-mother has an associative strength of .70). Therefore, Cho, Neely, Brennan, et al. (2017) and Coppens et al. (2016) proposed that the TE for target recall to the mediator could be coming from the mediator activating the cue, which in turn activated the target. If this is the case, the target would be recalled via the direct cue-target association, and not via a mediated chain. In both Cho, Neely, Brennan, et al. (2017) and Coppens et al. (2016), the cue-mediator forward association was strong for all pairs, as was so in Carpenter (2011). However, the pairs in one list also had a strong backward association from the mediator to the cue (as in mother-father-child), and in another list had a weak backward association from the mediator to cue (e.g., barracuda-fish-fast). The backward-chaining account predicts the TE for target recall to the mediator should be larger for the pairs with a strong mediator-to-cue backward association and perhaps even null for the pairs with a weak mediator-to-cue backward association from the mediator to the cue. Because the TEs for the two types of pairs were significant and equal in both studies, the backward-chaining account was not supported.

**Additional tests of the MEH**

Carpenter and Yeung (2017) conducted two experiments assessing the role of mediation with a slightly different procedure. They used weakly-related cue-target word pairs with varying strengths between the cues and the putative mediators (i.e., the primary associate to the cue) (e.g., weak association between the cue and mediator: ball-bat-boy; strong association between
the cue and mediator: milk-cow-water). Participants studied the cue-target word pairs three times, and then restudied half the word pairs, and were given a cued-recall test on the other half. The final cued-recall test was administered two days later, and participants were asked to recall the target to the cues. Carpenter and Yeung reasoned that if the MEH is correct, there should be a larger TE for the cue-target word pairs that have stronger cue-mediator strengths. Indeed, a larger TE emerged as the cue-mediator associative strength increased.

Across several experiments, Lehman and Karpicke (2016) searched for additional support of the MEH, but failed to find it. They used a lexical decision task to determine if mediators are activated during cue-target review testing. Participants studied cue-target word pairs (e.g., frame-portrait), and then completed a review phase in which they were tested on half of the word pairs and restudied the other half. In Experiments 1a and 1b, immediately after each individual review trial, either the associated mediator for the cue-target word pair (i.e., the cue’s primary associate, e.g., picture), an unrelated word to the cue-target word pair (e.g., shingle), or a pronounceable nonword (e.g., clett) was presented and a lexical decision was made to it, or nothing was presented. Experiment 2 used the same procedures except that the lexical decisions were made after all review trials had been presented. Lehman and Karpicke found that reaction times were faster for the mediators overall; however, this did not depend on whether the cue-target pair was tested versus restudied. Thus, there was no evidence that testing activated the mediators more than restudying.

A second goal of Lehman and Karpicke’s (2016) studies was to determine if generating mediators enhances target recall in the long run. They argued that if more mediators are generated, more pathways to target retrieval should be created. Therefore, generating more mediators should lead to better target recall on the final test. In four experiments, Lehman and
Karpicke first had participants study only the target words, and then in a second phase, generate associates to the cue words (as many as possible in Experiments 3a and 3b, two, four or six in Experiment 4, and two, three, four, five, or six in Experiment 5). Following a short distractor phase, participants completed a final test on target recall (cued-recall and free recall in Experiments 3a, 3b, and 4, and cued-recall only in Experiment 5). In all four experiments, there was no evidence that generating more mediators helped target recall on the final test. In fact, some of Lehman and Karpicke’s results suggested that generating fewer mediator associates enhanced final-test target recall. To account for this they offered an account whereby the activation of information semantically related to the target could actually suppress recall of the target by making it less distinctive. Furthermore, they suggested that the more mediators that are generated, the greater is the reduction in target distinctiveness.

**Strongest test of the MEH**

With the exception of Pyc and Rawson’s (2010) original experiment, none of the aforementioned experiments had participants recall both the mediator and the target to the cue. This procedure is needed to provide the most compelling test of the strongest version of the MEH, which says that target recall to the cue does not occur via a direct cue-target association but rather occurs only via the cue-mediator-target chain.

Cho, Neely, Brennan, et al. (2017) used Carpenter’s (2011) procedures and word pairs to examine spontaneous mediation, but, as did Pyc and Rawson (2010), they included a condition in which participants were given the cue and were asked to recall both the mediator and the target to it. Because participants were not asked to report a mediator and the mediator had not been presented, for mediator recall, the cue word was accompanied by the first letter of the cue’s primary associate, which is the putative mediator. The MEH predicts that target recall should
depend on mediator recall, and this dependency should be greater when the pairs had been tested rather than restudied during review. Disconfirming this prediction, Cho, Neely, Brennan, et al. (2017) found that target recall was slightly worse, not better, when the mediator was recalled vs. not recalled and that the TE was of the same magnitude whether the mediator was recalled vs. not recalled. Thus, their data did not support the MEH. However, it is possible that their results did not support the MEH because of the English word pairs that were used (e.g., mother-child). Although these previously used pairs were associatively unrelated, they were weakly related semantically unlike Pyc and Rawson’s (2010) completely unrelated Swahili-English pairs. Thus, mediation simply may not have been necessary for learning these word pairs.

The Mediator Shift Hypothesis

Pyc and Rawson (2012) proposed the mediator shift hypothesis (MSH) to complement the MEH. The MEH states that testing strengthens mediation in cue-target word pairs, and the MSH adds to it by suggesting that testing can strengthen mediation in part by prompting participants to shift to new, more effective mediators. Thus, a strength of the MSH is that it includes an explanation for how testing improves mediation relative to restudying. In a design similar to that of Pyc and Rawson (2010), Pyc and Rawson (2012) used Swahili-English word pairs (e.g., mshoni-tailor) and instructed participants to generate their own mediators that were phonologically similar to the cue and semantically related to the target (e.g., shoes) on an initial study phase and three additional review cycles. Participants either restudied or were tested on the word pairs in each review cycle, and were given the opportunity to generate new mediators each time. Two days later, all participants came back to the laboratory for a cued-recall final test, in which the target was to be recalled to the cue. After the cued-recall final test, the participants were presented either with the cue-target pair again (Experiment 1) or just the cue (Experiment
2) and were asked to recall their most recently generated mediator. Consistent with the MEH, Pyc and Rawson (2012) found a standard TE for cue-target recall.

Pyc and Rawson (2012) proposed that the TE for cue-target recall could be due to the test group shifting to more effective mediators during study/review. That is, if the participants who receive a review test fail to recall the target on a review phase cycle, they might shift to a more effective mediator on the next review cycle. In contrast, the participants who restudy the word pairs do not have a chance for retrieval failure during the review phases, which means they are not able to assess how effective their mediators are, and would then be less likely to switch to a new mediator on the next cycle. If the MSH is supported, Pyc and Rawson (2012) predicted: (1) a larger number of mediator shifts in the test group as compared to the restudy group, (2) a larger number of mediator shifts after a retrieval failure in the review phase as compared to retrieval success (which can only be assessed in the test group), and 3) higher mediator recall in the Session 2 final test for the test group as compared to the restudy group. This last prediction is consistent with the predictions of the MEH. All three predictions were confirmed in Pyc and Rawson’s (2012) Experiment 2.

**The Current Studies**

The current studies assessed the MEH and MSH with weakly semantically related English word pairs that should benefit from mediation more than the word pairs that Carpenter (2011) and Cho, Neely, Brennan, et al. (2017) used. Experiments 1 and 2 in the current paper evaluated the MEH under spontaneous mediation conditions using the most rigorous test of the MEH. That is, in the final test they had participants recall both the mediator and target to the cue to determine if (a) target recall would be very low when the mediator cannot be successfully retrieved and (b) the TE would be much larger when the mediator is successfully retrieved than
when it is not successfully retrieved. Experiments 3a and 3b assessed the MEH and MSH with explicit mediation instructions using the most rigorous test of the MEH and the same word pairs employed in Experiments 1 and 2.

In Experiment 1, semantically and associatively unrelated English cue-target word pairs (e.g., lion-stripes), which are analogous to Swahili-English word pairs, were used. The cue for each word pair had a strong primary associate that could be used as a mediator for the target, to which the mediator was associatively related (e.g., lion-tiger-stripes). In Experiment 2, the word pairs were English word pairs (e.g., now-milk) that were designed to be more similar to Pyc and Rawson’s (2010) word pairs in that the mediators were phonologically related to the cue and semantically related to the target (now-cow-milk). Both the semantic and phonological mediators would presumably aid the learning of these semantically and associatively unrelated cue-target pairs. In both experiments, participants studied the cue-target word pairs, then restudied or were tested on each, and then repeated the study(review cycles two more times. Experiments 1 and 2 had three study/review cycles, as in Pyc and Rawson (2010), rather than only one study/review cycle, as in Carpenter (2011) and Cho, Neely, Brennan, et al. (2017). Three study/review cycles provide the opportunity for test-potentiated learning (TPL)\(^1\) and/or for mediator shifting to occur. However, in the Cho, Neely, Brennan, et al. (2017) experiments, in which only one review test was given, there was no opportunity for participants in the test group to shift to a more effective mediator when they failed to successfully retrieve the target in the review test. Thus, Cho, Neely,

\(^1\) TPL refers to testing enhancing the encoding or learning of subsequently studied material (Arnold & McDermott, 2013; for reviews see Chan, Meissner, & Davis, 2018; Neely & Cho, 2014; Pastötter, & Bäuml, 2014; Yang, Chew, Sun, & Shanks, 2018). In one of the few studies (if not only study) to separate TPL from the beneficial effects of review testing on subsequent retrieval, Cho, Neely, Crocco, and Vitrano (2017) found that TPL and retrieval benefits from testing were of comparable magnitudes for both the review-tested pairs and pairs that were studied for the first time after the review test but were not themselves tested before the final test.
Brennan, et al.’s (2017) failure to support the MEH with spontaneous mediation and English word pairs could have been due to the absence of an opportunity for TPL and/or mediator shifting to occur.

In the current experiments, any mediation that occurs would be “spontaneous” in that participants were not required to report mediators either when the pairs were studied or were restudied or tested during review. Two days after the review, all participants completed a final test in which they were given each cue and asked to recall both the mediator and target to it, as in Pyc and Rawson (2010) and Cho, Neely, Brennan, et al. (2017). According to the MEH, target recall should depend on mediator recall, and this dependency should be greater when the pairs had been tested rather than restudied during review. To foreshadow, the results of Experiments 1 and 2 did not support the MEH.

Experiments 3a and 3b used the same materials as Experiments 1 and 2, respectively, but with explicit mediation instructions to determine if these materials are well suited for supporting the MEH when mediation is encouraged rather than being spontaneous. The procedure was modeled after Pyc and Rawson’s (2010) procedure. In both Experiments 3a and 3b, participants studied the cue-target word pairs and were asked to generate a mediator that was meaningfully related to both the cue and target (Experiment 3a) or that sounded like the cue and was related in meaning to the target (Experiment 3b). Following the initial study phase, the participants then either restudied or were tested on each cue-target pair and were again asked to generate a mediator. They were told that they could repeat the same mediator from the study phase or generate a new mediator. This study/review cycle was then repeated two more times, with the participants generating a mediator each time. Two days later, all participants completed a final test in which they were given each cue and asked to recall both the mediator and target to it, as in
the previous experiments. According to the MEH, target recall should depend on mediator recall, and this dependency should be greater when the pairs had been tested rather than restudied during review. Furthermore, according to the MSH, if test-enhanced mediation occurs, there should be more mediator switches for the test group as compared to the restudy group in the review cycles, more mediator switches following a target retrieval failure on the previous cycle as compared to a retrieval success, and higher mediator recall in Session 2 for the test group as compared to the restudy group.

**Experiment 1**

**Method**

**Design.** As shown in Figure 1, different groups of participants restudied or were given a cued-recall test with feedback on the cue-target pairs during the review phases. In the final test, both groups recalled both the mediator and target to each cue. This created a 2 (Type of Review: Test vs. Restudy) x 2 (Type of Final Test Recall: Mediator vs. Target) mixed-factor design, with Type of Review and Type of Final Test Recall as between- and within-participants variables, respectively.

**Materials.** Twenty-four mediator triads (cues-mediators-targets: e.g., lion-tiger-stripes) were taken from Balota and Lorch (1986). Forward associative strengths (FASs) and backward associative strengths (BASs) were measured using Nelson, McEvoy, and Schreiber’s (2004) free association norms (cue-mediator: $FAS_M = 0.27$, $BAS_M = 0.18$, mediator-target: $FAS_M = 0.13$, $BAS_M = 0.17$, cue-target: $FAS_M = 0.00$, $BAS_M = 0.00$). Semantic relatedness was measured using Landauer, Foltz, and Laham’s (1998) latent semantic analysis (LSA) norms (cue-mediator: $M = 0.52$; mediator-target: $M = 0.39$; cue-target: $M = 0.25$). All values for each triad are presented in Appendix A.
For the study and review phases, only the cues and targets were studied/tested as a word pair, without the mediator (e.g., lion-stripes). Therefore, the cue-target word pairs were studied and reviewed before the final test, whereas the putative mediators were not. For the final test, participants were asked to retrieve from each cue both the non-studied mediators, and the studied targets.

Participants. One hundred and thirty-one participants from the University at Albany, State University of New York participated in exchange for research credit in an introductory psychology course. None of the participants had previously participated in any other TE experiment in the lab. All participants reported normal or corrected-to-normal vision. The data from 11 participants (seven in the test group and four in the restudy group) were discarded because the participants failed to come back for Part 2 of the study. Of the remaining participants, sixty were randomly assigned to the test group, and sixty to the restudy group. The decision to test sixty participants in each of the test and restudy groups was based on that number of participants having been used in Experiment 2 of Cho, Neely, Brennan, et al. (2017), which was similar to the present experiment in that it also required participants to recall both the mediator and the target to each cue so as to provide the most compelling test of the strongest version of the MEH.

Procedure. Individually tested participants were told they would study lists of word pairs and receive a memory test, but they were not explicitly told how they would be tested. All pairs were presented in white Courier New font on a black background by a computer using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA).

After reading the instructions, all participants received Study List 1, for which each of the 24 cue-target pairs was presented for 2.5 s in a different randomized order for each participant.
Immediately after Study List 1, participants completed Review Phase 1, in which they either restudied the 24 cue-target pairs once or received one test with feedback on each of them. In the test group, each cue was presented in a different randomized order for each participant, and participants were given a maximum of 6 s to type its associated target. After 6 s, the correct answer was displayed for 1 s. In the restudy group, participants were told that they would be presented with the pairs again and that they should remember them for an upcoming memory test. The pairs were presented for 7 s in a different randomized order for each participant.

Immediately after all pairs had been tested or restudied in the first review phase, the participants began Study List 2. The additional study lists (Study List 2 and Study List 3), and the additional review phases (Review Phase 2 and Review Phase 3) were identical to Study List 1 and Review Phase 1, except that the pairs were presented in a different randomized order. Therefore, participants completed three cycles of study and review for the same 24 cue-target word pairs.

After all study and review phases were completed, the participants were given credit for the session and were asked to return to the laboratory 48 hours later, although they were not explicitly told what to expect when they returned for Session 2. Session 1 lasted approximately 20 minutes.

When the participants returned to the laboratory, they completed the cue-mediator-target cued-recall final test. Participants were now shown the same 24 cues from Session 1 in a different randomized order for each participant but were asked to retrieve two words for each cue: the putative mediator and the target. The participants were first provided the cue and were instructed to type a related word that comes to mind but was not studied during Session 1 (i.e., the mediator). Since the mediators were not studied in the experiment, the mediator’s first letter and underscores for each of its other letters were provided for the participants.
(e.g., lion-t_ _ _), following the procedure of Cho, Neely, Brennan, et al. (2017). After the participants typed the mediator, they pressed the “enter” key to advance to the next screen. The same cue was presented again but without any accompanying letter or underscores. This time, the participants were instructed to type the word that was paired with that cue in Session 1 of the experiment (i.e., the target). After the participants typed the associated target, they pressed “enter” to advance to the next trial with a new cue. They repeated this process for all 24 cues. Before beginning the final test, the participants were shown an example trial and asked if they had any questions. The final test was self-paced, and no feedback was given. Session 2 lasted approximately 10 minutes.

**Results**

**General statistical approach.** In all experiments, effect sizes for the $F$ statistics and independent sample $t$-tests are reported as partial eta-squared ($\eta_p^2$) and Cohen’s $d$, respectively. The number following the ± is the 95% confidence interval.

**Review test recall.** The percentages of mean correct target recall to the cues in Experiment 1 are shown in the first row of Table 1 for each of the three Session 1 review phases. As expected, review test recall increased significantly with every cycle of review (Review 1: 52 ± 5%, Review 2: 83 ± 4%, Review 3: 93 ± 2%), $F(2,118) = 318.818$, $p < .001$, $\eta_p^2 = .844$.

**Final test recall.** The first section of Table 2 presents the mean Session 2 final test correct recall percentages of both the mediators and targets to their cues.

A 2 (Type of Review: Test vs. Restudy) × 2 (Type of Final Test Recall: Mediator vs. Target) mixed-factor ANOVA was conducted, with the first factor being between-participants and the second factor, within-participants. Overall, mediator recall (86 ± 2%) was significantly higher than target recall (62 ± 4%), $F(1,118) = 138.656$, $p < .001$, $\eta_p^2 = .540$, which is not
surprising, given that the cue-target pairs were unrelated, and the mediators were primary associates of the cue. Furthermore, the first letter of each mediator was given as a prompt, which could also explain the high mediator recall. Also, averaged across mediator and target recall, there was a robust 18 ± 5% TE, as indicated by the significant main effect of Type of Review, $F(1,118) = 45.410, p < .001, \eta_p^2 = .278$.

The interaction between Type of Review and Type of Final Test Recall was also significant, $F(1,118) = 67.668, p < .001, \eta_p^2 = .364$, due to a 34 ± 8% larger TE for target recall than mediator recall. As expected, the 35% TE for target recall was significant, $F(1,118) = 64.638, p < .001, \eta_p^2 = .354$. However, the 1% TE for mediator recall was not significant, $F(1,118) = .240, p = .625, \eta_p^2 = .002$. This is incongruent with the strongest version of the MEH, which says that testing strengthens both the cue-to-mediator and the mediator-to-target associations.

The first section of Table 3 presents the percentages of mean target recall in the Session 2 final test in Experiment 1 conditionalized on recall of the corresponding mediator. A 2 (Type of Review: Test vs. Restudy) × 2 (Mediator Recall: Mediator recalled vs. Mediator not recalled) mixed-factor ANOVA was conducted, with the first factor being between-participants and the second, within-participants. As discussed earlier, this analysis provides a more rigorous test of the strongest version of the MEH. Once again, the Type of Review Test main effect was significant, $F(1,107) = 38.221, p < .001, \eta_p^2 = .26$, based on the 33 ± 9% TE. Additionally, target recall to the cue was 8 ± 6% greater overall when the mediator could be recalled (62 ± 5%) as compared to when it could not be recalled (54 ± 7%), $F(1,107) = 7.237, p = .008, \eta_p^2 = .063$.

---

2 Eleven participants (five in the test group and six in the restudy group) were excluded from the analysis because they recalled all of the mediators during Session 2.
which is congruent with spontaneous mediation. However, contrary to the MEH, the TE for target recall was not larger when the mediator was recalled than when it was not recalled, $F(1,107) = .725, p = .396, \eta^2_p = .007$, as shown by the nonsignificant $5 \pm 11\%$ difference in TEs for target recall when the mediator could be recalled as compared to when it could not, $t(107) = .851, p = .396, d = .163$.

To further assess the nonsignificant $5 \pm 11\%$ interaction, a Bayes factor analysis (e.g., Rouder, Morey, Speckman, & Province, 2012) was conducted using JASP (Version 0.11.1; JASP Team, 2019) to compare a model with Mediator Recall, Type of Review and the interaction all included and a model with only the main effects of Mediator Recall and Type of Review included. This produced a Bayes factor of 3.79, demonstrating that the data are 3.79 times more likely under the model with the two main effects than the model with the two main effects plus the interaction. This can be considered moderate evidence for a model without the interaction (Van Doorn et al., 2019). Thus, contrary to the MEH, the combined results suggest that testing does not enhance spontaneous mediation for the weakly-related English word pairs used in Experiment 1.

**Experiment 2**

The word pairs used in Experiment 1 were chosen because a mediation strategy should enhance the participant’s ability to learn them. However, both the cue-to-mediator links and the mediator-to-target links were associatively related pre-experimentally in Experiment 1, unlike in Pyc and Rawson (2010). Experiment 2 was conducted with word pairs that were more similar to
the word pairs used in Pyc and Rawson (2010) in that the cue-to-mediator relation was phonologically based, and the mediator and target were associatively related pre-experimentally.\footnote{3}

**Method**

Unless explicitly noted otherwise, the design, participant characteristics and the procedures were identical to those of Experiment 1.

**Materials.** Twenty-four mediator triads (cues-mediators-targets: e.g., now-cow-milk) were created from the Balota and Lorch (1986) triads. Specifically, the mediator-target links were kept identical to the Balota and Lorch triads, but the cue was changed to rhyme with the mediator. For example, one of the original triads from Balota and Lorch (1986) was bull-cow-milk, whereas in Experiment 2, the triad was changed to now-cow-milk to be more analogous to the word pairs used in Pyc and Rawson (2010). Forward associative strengths (FASs) and backward associative strengths (BASs) were measured using Nelson et al.’s (2004) free association norms (cue-mediator: FAS $M = 0.00$, BAS $M = 0.00$, mediator-target: FAS $M = 0.13$, BAS $M = 0.17$, cue-target: FAS $M = 0.00$, BAS $M = 0.00$). Semantic relatedness was measured using Landauer et al.’s (1998) latent semantic analysis (LSA) norms (cue-mediator: $M = 0.21$; mediator-target: $M = 0.39$; cue-target: $M = 0.22$). For the mediators, the probability of producing the mediator from the ending sound (ESGQ) was also measured using Nelson et al.’s (2004) free association norms (ESGQ $M = .16$). For example, the ESGQ value of .12 for the mediator cow (in the triad now-cow-milk), indicates the probability of producing cow from the ending sound “ow”. All values for each triad are presented in Appendix B.

\footnote{3 The cue-to-mediator relation in Experiment 2 was phonologically based. Additionally, half of the cue-to-mediator materials were also orthographically related (e.g., now-cow are phonologically related and are orthographic neighbors). Thus, half of the mediators confounded phonology and orthography. However, this confound also existed in Pyc and Rawson (2010).}
Participants. There were one hundred and fifty-four participants. The data from 34 participants total were discarded: 26 participants (nine in the test group and 17 in the restudy group) failed to come back for Part 2 of the study, seven (five in the test group and two in the restudy group) failed to follow directions, and one (in the test group) did not have normal or corrected-to-normal vision. All remaining participants reported normal or corrected-to-normal vision. Of the remaining participants, sixty were randomly assigned to the test group, and sixty to the restudy group.

Results

Review test recall. The percentages of mean correct target recall to the cues in Experiment 2 are shown in the second row of Table 1 for each of the three Session 1 review phases. Review test recall was significantly lower than in Experiment 1 but followed a similar pattern. That is, review test recall increased significantly with every cycle of review, $F(2,118) = 254.275$, $p < .001$, $\eta_p^2 = .812$. The lower review test recall in Experiment 2 could be due to the change in the cue-mediator links for Experiment 2 having decreased the associative strength between the cues and mediators, as compared to the associative strengths between the cues and mediators in Experiment 1 (Experiment 1: FAS $M = 0.27$, BAS $M = 0.18$; Experiment 2: FAS $M = 0.00$, BAS $M = 0.00$). The higher review test recall in Experiment 1 suggests that spontaneous mediation occurred, and is congruent with the Experiment 1 finding that final test target recall was significantly higher when the mediator could be recalled versus when it could not be recalled.

Final test recall. The second section of Table 2 presents the mean Session 2 final test correct recall percentages of both the mediators and targets to their cues. A $2 \times 2$ (Type of Review: Test vs. Restudy) $\times 2$ (Type of Final Test Recall: Mediator vs. Target) mixed-factor ANOVA
was conducted, with the first factor being between-participants and the second factor, within-participants. Overall, there was no difference between mediator recall (33 ± 4%) and target recall (37 ± 4%), $F(1,118) = 3.357, p = .127, \eta^2_p = .020$, which is in contrast to Experiment 1, in which mediator recall was significantly higher than target recall. As discussed above, this is likely because in Experiment 1, the cues and mediators were more strongly associated as compared to the cues and mediators in Experiment 2. Also, averaged across mediator and target recall, there was a robust 10 ± 7% TE, as indicated by the significant main effect of Type of Review, $F(1,118) = 8.076, p = .005, \eta^2_p = .064$.

The interaction between Type of Review and Type of Final Test Recall was also significant, $F(1,118) = 28.113, p < .001, \eta^2_p = .192$, due to a 23 ± 9% larger TE for target recall than mediator recall. As expected, the 21% TE for target recall was significant, $F(1,118) = 28.690, p < .001, \eta^2_p = .196$. However, the -2% TE for mediator recall was not significant, $F(1,118) = .330, p = .567, \eta^2_p = .003$. These results replicate Experiment 1, and are again incongruent with the strongest version of the MEH.

The second section of Table 3 presents the percentages of mean target recall in the Session 2 final test in Experiment 2 conditionalized on recall of the corresponding mediator. A 2 (Type of Review: test vs. restudy) × 2 (Mediator Recall: Mediator recalled vs. not recalled) mixed-factor ANOVA was conducted for target recall, with the first factor being between-participants and the second, within-participants. Once again, the Type of Review Test main effect was significant, $F(1,117) = 28.646, p < .001, \eta^2_p = .197$, based on the 22 ± 8% TE. Additionally, target recall to the cue was 24 ± 5% overall greater when the mediator could be

---

4 One participant in the restudy group was excluded from the analysis for recalling all of the mediators during Session 2.
recalled (54 ± 6%) as compared to when it could not be recalled (30 ± 4%), $F(1,117) = 100.024$, $p < .001$, $\eta^2_p = .461$, which is congruent with spontaneous mediation. However, contrary to the MEH, the TE for target recall was again not larger when the mediator was recalled than when it was not recalled, $F(1,117) = 1.267$, $p = .263$, $\eta^2_p = .011$, as shown by the nonsignificant 5 ± 10% difference in TEs for target recall when the mediator could be recalled as compared to when it could not, $t(117) = 1.126$, $p = .263$, $d = .206$. The results suggest that testing did not enhance spontaneous mediation for the Experiment 2 word pairs either. It is also noteworthy that these two failures to find that the TE is larger when the mediator is successfully recalled than when it is not recalled generalize across substantively different levels of recall in the final tests of Experiments 1 and 2.

To further assess the nonsignificant 5 ± 10% interaction, a Bayes factor analysis was computed in the same way as Experiment 1 to compare a model with Mediator Recall, Type of Review and the interaction all included and a model with only the main effects of Mediator Recall and Type of Review included. This produced a Bayes factor of 3.02, demonstrating that the data are 3.02 times more likely under the model with the two main effects than the model with the two main effects plus the interaction. Consistent with Experiment 1, this can be classified as moderate evidence for a model without the interaction. However, it must be acknowledged that in both Experiments 1 and 2, the BF values were close to the value of 3 needed to rise above anecdotal evidence (Van Doorn et al., 2019).

**Experiments 1 and 2.** The data from both Experiments 1 and 2 showed that the TE was 5 ± 8% greater when the mediator was successfully recalled than when it was not. Because this difference **numerically** (but not statistically, $p = .180$) supports the MEH and because Experiments 1 and 2 differed only in the English word pairs used as materials, the data from
Experiments 1 and 2 were combined to increase statistical power. When the data from Experiments 1 and 2 were combined, target recall to the cue was $16 \pm 4\%$ higher when the mediator could be recalled ($58 \pm 4\%$) as compared to when it could not be recalled ($42 \pm 4\%$), $F(1,224) = 73.698, p < .001, \eta_p^2 = .248$, congruent with the MEH. However, contrary to the MEH, this difference was not enhanced by testing. The TE for target recall was not larger when the mediator could be recalled than when it could not be recalled, $F(1,224) = 1.920, p = .167, \eta_p^2 = .008$, as shown by the nonsignificant $5 \pm 18\%$ difference in TEs for target recall when the mediator could be recalled as compared to when it could not, $t(117) = 1.337, p = .183, d = .177$. However, the computed Bayes Factor for the combined data only weakly support the claim that the data fail to support the MEH in that it shows that the data are only $3.12$ times more likely under a model with Type of Review and Mediator Recall as compared to a model that also includes the interaction. Per Van Doorn et al. (2019), this $3.12$ value is barely above the threshold value of $3$ needed to call the evidence for there being no interaction moderately strong (as opposed to anecdotal). Nevertheless, the overall weight of the evidence fails to support testing-enhanced spontaneous mediation with English word pairs.

**Experiments 3a and 3b**

The results of Experiments 1 and 2 did not favor the MEH when English word pairs were learned and no instructions were given for participants to use a mediator to learn them. Without such instructions, spontaneous mediation was being assessed. The aim of Experiment 3 was to examine whether the failure to support the MEH in Experiments 1 and 2 was due to (a) English word pairs rather than Swahili-English pairs being learned and/or (b) the absence of instructions that told participants to employ a mediation strategy to learn these pairs. To evaluate this, Experiments 3a and 3b used the same word pairs as those used in Experiments 1 and 2,
respectively, but with procedures that closely followed those of Pyc and Rawson (2010). That is, participants studied the pairs once and then received three consecutive review cycles. During the reviews, one group of participants restudied the pairs and another group was given a cued-recall test on each pair with the correct pair being presented immediately afterward as feedback. Most important, as in Pyc and Rawson (2010) but unlike in Experiments 1 and 2, on each study and review trial, participants were asked to generate and type their own mediator for each word pair. In Experiment 3a, they were instructed to generate a mediator that is semantically associated to both the cue and target (e.g., “doctor” for nurse-lawyer); in Experiment 3b, they were instructed to generate a mediator that is phonologically related to the cue and also has a meaning that is associated to the target (e.g., “nose” for toes-smell).

Although the general procedures of Pyc and Rawson (2010) were followed, there were three differences between Experiments 3a and 3b and Pyc and Rawson’s (2010) experiment: (1) They used English word pairs whereas Pyc and Rawson (2010) used Swahili-English word pairs. (2) In order to increase the likelihood that participants would sign-up and come back for the second session, the current experiments used a two-day retention interval between the first and second sessions rather than Pyc and Rawson’s (2010) one-week retention interval. (The two-day retention interval is consistent with that used by Pyc & Rawson, 2012.) (3) Whereas Pyc and Rawson (2010) gave participants one of three different final tests during Session 2, all participants in Experiments 3a and 3b were given the same final test in Session 2. They were presented with each previously studied cue and asked to respond with the last mediator they had generated to it and to then recall its associated target. In the two other final tests in Pyc and Rawson (2010), different groups of participants either received (a) the cue alone or (b) the cue and the mediator that had been most recently generated to it. Both groups were to report the
cue’s associated target. These groups were not included in Experiment 3 because they are not necessary for affording the most compelling assessment of the validity of the strongest version of the MEH.

The generation of mediators over several study/review cycles allows for the predictions of the MSH to be tested in addition to the MEH. That is, the MSH predicts that (1) the test group will shift mediators more often than the restudy group, and (2) that mediator shifts will be more common after a previous failure to recall the target as compared to after successful recall of the target. The third prediction of the MSH is also a prediction of the MEH, and states that the test group will recall more mediators at final test than the restudy group.

**Experiment 3a**

**Method.**

**Design.** As shown in Figure 2, different groups of participants restudied or were given a cued-recall test on the cue-target pairs during the review phases. In the final test, both groups recalled both the mediator and target to each cue. This created a 2 (Type of Review: Test vs. Restudy) x 2 (Type of Final Test Recall: Mediator vs. Target) mixed-factor design, with Type of Review and Type of Final Test Recall as between- and within-participants variables, respectively.

**Materials.** The cue-target word pairs from Experiment 1 were used.

**Participants.** Fifty-three undergraduate students having the same characteristics as the participants in Experiments 1 and 2 participated in the experiment. The same criteria used for discarding data in Experiments 1 and 2 was employed. The data from five participants total were discarded: four participants in the test group failed to come back for Part 2 of the study, and one participant in the test group failed to follow directions. All remaining participants reported
normal or corrected-to-normal vision. Of the remaining participants, twenty-four were randomly assigned to the test group, and twenty-four to the restudy group. This number of participants is comparable to the 19-20 participants that were tested in Pyc and Rawson’s (2010) test and restudy groups.

**Procedure.** Individually tested participants were told they will study lists of word pairs and receive a memory test, but they were not explicitly told how they would be tested. All pairs were presented in the same way as in Experiments 1 and 2.

At the start of the experiment, participants read instructions stating that they would first study a list of cue-target word pairs and that they should generate and type their own mediators for each of the pairs. As in Pyc and Rawson (2010), participants were given explicit instructions on how to generate their own mediators. That is, they read instructions stating that they should generate a mediator related in meaning to both the cue and target, and they were given an example (i.e., “tiger” for lion-stripes). After reading the instructions, all participants began Study List 1, in which each of the 24 cue-target pairs were presented with a textbox in which participants were to type their generated mediators. The word pairs were presented in a different randomized order for each participant, and all trials were self-paced. Immediately after Study List 1, the participants completed Review Phase 1, in which they either restudied the 24 cue-target pairs once (restudy group) or received one cued-recall test on each pair immediately followed by the correct cue-target pair again (test group). In each trial in the review phase, participants were again given a textbox in which to type their generated mediators. Participants were told they could use the same mediators from the study phase or switch to a new mediator if they felt that would be useful. All pairs were presented in a different randomized order for each participant, and all trials were self-paced, similar to the study phase. Immediately after Review
Phase 1, participants completed the same review phase two more times. The additional review phases were identical to Review Phase 1, except that the pairs were presented in a different randomized order each time. Therefore, as in Pyc and Rawson (2010), participants completed one study cycle, and three review cycles for the same 24 cue-target word pairs. After all study and review phases were completed, participants were given credit for the session and were asked to return to the laboratory 48 hours later. Session 1 lasted approximately 20 minutes.

When the participants returned to the laboratory, they completed a cued-recall final test. Participants were shown the same 24 cues from Session 1 in a newly randomized order and were asked to retrieve two words for each cue: their last generated mediator and the studied target. In contrast to Experiments 1 and 2, because the participants now reported the mediators they were using, they were not prompted with the first letter of the mediator in the final test. This lack of constraint on the mediator that could be recalled in the final test allows a better test of mediator use and the MEH. All trials were self-paced, and participants pressed “enter” to advance to the next trial. After all 24 trials had been given, the experiment ended, and all participants were given credit for the session. Session 2 lasted approximately 10 minutes.

Results

Review test recall. The percentages of mean correct target recall to the cues in Experiment 3a are shown in the third row of Table 1 for each of the three Session 1 review phases. A one-way ANOVA was conducted on the three review phases. As expected, review test recall increased significantly with every cycle of review (Review 1: 72 ± 8%, Review 2: 87 ± 3%, Review 3: 94 ± 3%), $F(2,46) = 42.245, p < .001, \eta^2_p = .647$.

“Specified” Mediators. The top half of Table 4 displays the mean percentage of times the “specified” mediator was used in Experiment 3a, with “specified” mediator being defined as
the mediator specified in the study materials from Balota and Lorch (1986) (i.e., a strong associate of the cue that is also associated to the target, as is the mediator “tiger” for the cue-target pair lion-stripes). Averaged across the study phase, the three review phases and across the Test and Restudy groups, participants generated and reported “specified” mediators 48 ± 5% of the time. As shown by the 95% confidence intervals in the top half of Table 4, in no case was there a significant difference in “specified” mediator generation between the test and restudy groups.

The top half of Table 5 presents the total percentage of trials with mediator shifts, broken down by whether the shift was from a “specified” mediator or an “unspecified” mediator. An “unspecified” mediator was any mediator that was not the “specified” mediator (e.g., generating “animal” for lion-stripes, instead of the “specified” mediator “tiger”). Following Pyc and Rawson (2012), a mediator shift was defined as generating a mediator that was different from the mediator generated in the immediately prior study or review cycle. A 3 (Review Phase: Review 1, Review 2, Review 3) x 2 (Mediator Shift: from a “specified” mediator vs. from an “unspecified” mediator) x 2 (Type of Review: Test vs. Restudy) mixed-factor ANOVA was conducted with the first two factors being within-participants and the second, between-participants. Overall, there was a larger percentage of trials with mediator shifts from an “unspecified” mediator (9 ± 2%) as compared to a “specified” mediator (1 ± 1%). This suggests that even when target recall had failed when the “specified” had been used on the immediately preceding trial, on the next trial they realized that it would likely be a more useful mediator than any other mediator they could think of. Furthermore, as shown by the parenthesized values for shifts from “unspecified” mediators, there was a higher percentage of shifts from an “unspecified” mediator to another “unspecified” mediator than from an “unspecified” mediator.
to a “specified” mediator. This was likely due to there being more potential “unspecified” mediators than “specified” mediators. There was also a significant interaction between mediator shifts and type of review, demonstrating a larger 6 ± 4% TE for a shift from an “unspecified” mediator as compared to a shift from a “specified” mediator. The data show that participants are using the “specified” mediators and are less likely to shift mediators after using the “specified” mediators. This suggests that the “specified” mediators were in fact more useful mediators than were “unspecified” mediators.

MSH Predictions. Following the lead of Pyc and Rawson (2012), the top half of Table 6 presents for Experiment 3a the total percentages of trials with mediator shifts, broken down by the test versus restudy group and whether or not the target was successfully recalled to its cue in the immediately preceding phase. The first section of the top half of Table 6 presents the mean percentages of trials with mediator shifts for the test and restudy groups over the three phases. To examine whether the test group shifted mediators more often than the restudy group (i.e., prediction 1 of the MSH, which was confirmed in Pyc & Rawson, 2012), a 3 (Phase: Review 1, Review 2, Review 3) x 2 (Type of Review: Test vs. Restudy) mixed-factor ANOVA was conducted on the percentage of trials with mediator shifts, with the first factor being within-participants and the second, between-participants. Consistent with the MSH, the (7 ± 6%) higher mean percentage of trials with mediator shifts for the test group (14 ± 4%) than for the restudy group (7 ± 4%) was statistically significant, $F(1,46) = 6.317, p = .016, \eta^2_p = .121$. The main effect of phase was also significant, $F(2,46) = 9.824, p < .001, \eta^2_p = .176$, showing a decrease in mediator shifts across phases. Specifically, averaged over the test and restudy groups, the 6 ± 4% greater percentage of trials with mediator shifts on Review 1 (15 ± 4%) as compared to Review 2 (9 ± 4%) was significant ($p < .001$), but the (1 ± 4%) difference in the percentage of trials with
mediator shifts on Review 2 (9 ± 4%) as compared to Review 3 (8 ± 2%) was not significant ($p = .389), F(2,46) = .040, p = .960, \eta^2_p = .001. The interaction between Review Phase and Type of Review was not significant, F(2,46) = .040, p = .960, \eta^2_p = .001.

The second section of Table 6 presents the mean percentage of trials with mediator shifts following target recall failure or success on the immediately preceding review phase. To assess Pyc and Rawson’s (2012) prediction that mediator shifts are more likely to occur after target recall failure (i.e., prediction 2 of the MSH), a 2 (Review Phase: Review 2 vs. Review 3) x 2 (Previous Review Recall: Success vs. Failure) repeated-measures ANOVA was conducted on the percentage of trials with mediator shifts. In contrast to the MSH’s prediction, mediator shifts were equally likely to occur following recall failure (5 ± 3%) and recall success (7 ± 3%), $F(1,23) = 1.396, p = .249, \eta^2_p = .057. The interaction was significant, demonstrating the nonsignificant 1 ± 5% greater percentage of mediator shifts following target recall failure versus target recall success in Review 2 was significantly different from the significant 5 ± 4% smaller percentage of mediator shifts following target recall failure versus target recall success in Review 3, $F(1,23) = 13.963, p < .001, \eta^2_p = .378. It is important to note that the latter 5 ± 4% difference was in the direction opposite to that predicted by the MSH. That is, the percentage of trials with mediators shifts was larger in Review Phase 3 following target recall success (8 ± 4%) as compared to target recall failure (3 ± 1%). It is not clear why this highly anomalous result emerged, but it is clear that the results do not support prediction 2 of the MSH. Although the test group engaged in more mediator shifting during the review phases as predicted by the MSH, it did not seem to depend on previous target recall failures versus successes, contrary to the MSH.

**Final test recall.** The last column of the top half of Table 4 presents the mean percentage of “specified” mediators that were recalled for the final test. As noted earlier, “specified”
mediators are the mediators that Balota and Lorch (1986) specified in their materials. Each specified mediator was a strong associate of the cue and was also associated to the target (e.g., “tiger” for lion-stripes). The 5 ± 10% difference in “specified” mediator recall between the test (46 ± 9%) and restudy groups (41 ± 6%) was not significant ($p = .323$).

The third section of Table 2 presents the mean Session 2 final test correct recall percentages of a target to its cue and of the mediator that was reported in Review Phase 3 for that same cue. Similar to the previous experiments, a 2 (Type of Review: Test vs. Restudy) × 2 (Type of Final Test Recall: Mediator vs. Target) mixed-factor ANOVA was conducted, with the first factor being between-participants and the second factor, within-participants. Consistent with the Session 2 instructions, correct mediator recall was scored using the participant’s most recently generated mediator (i.e., the mediator from Review 3). As in Experiment 1, overall mediator recall (74 ± 5%) was significantly higher than target recall (64 ± 5%), $F(1,46) = 29.098$, $p < .001$, $\eta^2_p = .387$. Also, averaged across mediator and target recall, there was a robust 21 ± 10% TE, as indicated by the significant main effect of Type of Review, $F(1,46) = 19.556$, $p < .001$, $\eta^2_p = .298$.

As in Experiment 1, which used the same pairs, the interaction between Type of Review and Type of Final Test Recall was also significant, $F(1,46) = 70.751$, $p < .001$, $\eta^2_p = .606$, due to a larger TE for target recall than mediator recall. As expected, the 37 ± 10% TE for target recall was significant, $F(1,46) = 56.057$, $p < .001$, $\eta^2_p = .549$. However, the 5 ± 11% TE for mediator recall was not significant, $F(1,46) = .953$, $p = .334$, $\eta^2_p = .020$. The nonsignificant TE for

---

5 The 2 (Type of Review: Test vs. Restudy) × 2 (Type of Final Test Recall: Mediator vs. Target) mixed-factor ANOVA was repeated to assess if participants were recalling mediators that were not their most recently generated mediators. Mediator recall remained similar when the participants’ Review 1 mediators were considered as correct mediator recall (73 ± 5%) and when Review 2 mediators were considered as correct mediator recall (73 ± 5%).
mediator recall fails to confirm both the MEH and prediction 3 of the MSH, but is consistent with the TEs obtained for Experiments 1 and 2. Compared to the restudy group, the test group engaged in more mediator switching during the review phases, but this did not translate to recalling more mediators in Session 2 two days later.

The third section of Table 3 presents the mean percentages of target recall in Session 2, conditionalized on whether or not the participants recalled the corresponding mediator. A 2 (Type of Review: Test vs. Restudy) × 2 (Mediator Recall: Mediator recalled vs. Mediator not recalled) mixed-factor ANOVA was conducted, with the first factor being between-participants and the second, within-participants.\\(^6\\) Similar to the results of the previous experiments, the Type of Review Test main effect was significant, $F(1,44) = 41.330, p < .001, \eta_p^2 = .484$, based on the 32 ± 10% TE. Additionally, target recall to the cue was 40 ± 10% overall greater when the mediator could be recalled (71 ± 6%) as compared to when it could not be recalled (31 ± 8%), $F(1,44) = 63.338, p < .001, \eta_p^2 = .590$, which is congruent with mediation. However, contrary to the MEH, the TE for target recall was not larger when the mediator was recalled than when it was not recalled, $F(1,44) = .118, p = .732, \eta_p^2 = .003$, as shown by the nonsignificant 3 ± 20% difference in TEs for target recall when the mediator could be recalled as compared to when it could not, $t(44) = .344, p = .732, d = .10$.\\(^7\\) Consistent with the previous experiments, a computed Bayes Factor supported this result. The data are 2.91 times more likely

---

\\(^6\\) Two participants in the test group were excluded from the analysis because they recalled all of the mediators during Session 2.
\\(^7\\) The same pattern of results emerged when considering mediators recalled from Review 1 and 2, instead of the most recently generated mediator. There was a nonsignificant 3 ± 17% difference in TEs for target recall when the mediator from Review 1 could be recalled in Session 2 as compared to when it could not, and a nonsignificant 0 ± 19% difference in TEs for target recall when the mediator from Review 2 could be recalled in Session 2 as compared to when it could not.
under a model with only Type of Review and Mediator Recall main effects as compared to a model that includes the interaction. Per Van Doorn et al. (2019), this 2.91 value means the data provide only anecdotal evidence for there being no interaction. The results replicate the results of Experiments 1 and 2 and suggest that the MEH/MSH is not supported with the current English word pairs even when explicit mediation instructions are given.

**Experiments 1 and 3a**

To assess the effects of explicit mediation instructions with semantically/associatively related cues and mediators, a 2 (Experiment: 1 vs. 3a) × 2 (Type of Review: Test vs. Restudy) × 2 (Mediator Recall: Mediator recalled vs. not recalled) mixed-factor ANOVA was conducted on target recall, to compare performance on the two experiments that used materials with semantically related cues and mediators. When the data from Experiments 1 and 3a were combined, the overall 32 ± 9% TE was significant, with the overall 71 ± 6% target recall in test group being greater than the 38 ± 6% recall in the restudy group, $F(1,151) = 53.848, p < .001, \eta^2_p = .263$. Furthermore, target recall to the cue was overall 24 ± 5% higher when the mediator could be recalled (66 ± 4%) as compared to when the mediator could not be recalled (43 ± 6%), $F(1,151) = 75.510, p < .001, \eta^2_p = .333$, congruent with mediation. However, contrary to the MEH, this difference was not enhanced by testing. The TE for target recall was not larger when the mediator could be recalled as compared to when it could not be recalled, $F(1,151) = .568, p = .452, \eta^2_p = .004$, as shown by the nonsignificant 4 ± 11% difference in TEs for target recall when the mediator could be recalled as compared to when it could not, $t(153) = .678, p = .498, d = .108$. This conclusion is also supported by the computed Bayes Factor. The data are 4.96 times more likely under a model with only Type of Review and Mediator Recall main effects as compared to a model that also includes the interaction. This Bayes Factor value is higher than the...
previous Bayes-Factor values which were close to 3.00 but the data still provide only moderate support for there being a null interaction per Van Doorn et al. (2019). Furthermore, there was a significant interaction showing that the difference in target recall when the mediator could be recalled versus when the mediator could not be recalled was larger for Experiment 3a than for Experiment 1, $F(1,151) = 34.553, p < .001, \eta_p^2 = .186$. This likely was due to the difference in instructions between the two experiments. The participants in Experiment 3a were explicitly told to use mediation, whereas any mediation that would occur in Experiment 1 would be spontaneous. Importantly, the three-way interaction was not significant, $F(1,151) = .019, p = .892, \eta_p^2 = .000$, demonstrating that there is no difference in testing-enhanced semantic/associative mediation when participants are explicitly required to report mediators as compared to when they are not. A computed Bayes Factor shows that the best model is a model with main effects of Experiment, Type of Review, and Mediator Recall, and an interaction between Experiment and Mediator Recall. This model is 75.89 times more likely than a model containing the three-way interaction. Per Van Doorn et al. (2019) these data provide strong support for the hypothesis that the requirement that the participant report a word (the mediator) that can be used to link up the cue and target leads to greater recall 2 days later of the word that was reported as the mediator. However, this enhancement of mediator recall when participants were required to overtly report the mediator was the same whether that enforced report occurred during a review that involved testing or involved restudying.

**Experiment 3b**

**Method.**

**Design.** Unless explicitly noted otherwise, the design, participant characteristics, and the procedures were identical to those of Experiment 3a.
Materials. The cue-target word pairs from Experiment 2 were used.

Participants. The data from twenty-two participants total were discarded: thirteen participants (three in the test group and ten in the restudy group) failed to come back for Part 2 of the study, six (five in the test group and one in the restudy group) failed to follow directions, one participant in the test group did not have normal or corrected-to-normal vision, and two participants (one in the test group and one in the restudy group) experienced a computer error. All remaining participants reported normal or corrected-to-normal vision. Of the remaining participants, twenty-four were randomly assigned to the test group, and twenty-four to the restudy group.

Procedure. The procedure was the same as in Experiment 3a, except for the explicit instructions on how participants should generate their own mediators. That is, for Experiment 3b, participants read instructions stating that they should generate a mediator that sounds like the cue and is semantically/associatively related to the target, and they were given the example “bread” for dread-butter.

Results

Review test recall. The percentages of mean correct target recall to the cues in Experiment 3b are shown in the fourth row of Table 1 for each of the three Session 1 review phases. As expected, review test recall increased significantly with every cycle of review (Review 1: 42 ± 9%, Review 2: 61 ± 9%, Review 3: 71 ± 9%), \( F(2,46) = 76.933, p < .001, \eta^2_p = .770. \)

“Specified” Mediators. The bottom half of Table 4 provides the mean percentage of times the “specified” mediator was used in Experiment 3b, with “specified” mediator defined as a word phonologically related to the cue and semantically related to the target (e.g., “ring” for
bring-finger). Averaged across the study phase, the three review phases and the final test and across the Test and Restudy groups, participants generated and reported “specified” mediators 45 ± 8% of the time, which is highly similar to the 48% overall mean observed in Experiment 3a. It should be noted that ten of the items had two-three other mediators other than the specified mediator that could fit within the study instructions (e.g., the specified mediator for ache-pie was “cake”, but some participants used “bake” instead”). For consistency, these other appropriate mediators were not included as “specified” mediators in computing the means shown in Table 4. As shown by the 95% confidence intervals in the bottom half of Table 4, in no case was there a significant difference in “specified” mediator generation between the test and restudy groups.

The bottom half of Table 5 presents the total percentage of trials with mediator shifts, broken down by whether the shift was from a “specified” mediator or an “unspecified” mediator. An “unspecified” mediator was any mediator that was not a “specified” mediator (e.g., generating “string” for bring-finger, instead of the “specified” mediator of “ring”). A 3 (Review Phase: Review 1, Review 2, Review 3) x 2 (Mediator Shift: From a “specified” mediator vs. from an “unspecified” mediator) x 2 (Type of Review: Test vs. Restudy) mixed-factor ANOVA was conducted with the first two factors being within-participants and the second, between-participants. As in Experiment 3a, overall there was a larger percentage of trials with mediator shifts from an “unspecified” mediator (16 ± 5%) as compared to a “specified” mediator (2 ± 1%). Once again, this suggests that even when target recall had failed when the “specified” had been used on the immediately preceding trial, on the next trial they realized that it would likely be a more useful mediator than any other mediator they could think of. Furthermore, as was also observed in Experiment 3a, there was a higher percentage of shifts from an “unspecified” mediator to another “unspecified” mediator than from an “unspecified” mediator to a “specified”
mediator (as shown by the parenthesized values for shifts from “unspecified” mediators). Once again, this is likely due to there being more potential “unspecified” mediators than “specified” mediators. There was also a significant interaction between mediator shifts and type of review, demonstrating a larger $7 \pm 9\%$ TE for a shift from an “unspecified” mediator as compared to a shift from a “specified” mediator. Similar to Experiment 3a, this suggests that the “specified” mediators were indeed useful mediators for the participants.

**MSH Predictions.** The bottom half of Table 6 presents for Experiment 3b the total percentages of trials with mediator shifts, broken down by the test versus restudy group, and whether or not the target was successfully recalled to its cue in the immediately preceding phase. The first section of the top half of Table 6 presents the mean percentages of trials with mediator shifts for the test and restudy groups over the three phases. To examine if the test group shifted mediators more during the review phases than the restudy group (i.e., prediction 1 of the MSH, which was confirmed in Pyc & Rawson, 2012), a 3 (Review Phase: Review 1, Review 2, Review 3) x 2 (Type of Review: Test vs. Restudy) mixed-factor ANOVA was conducted on the percentage of trials with mediator shifts, with the first factor being within-participants and the second factor, between-participants. Consistent with prediction 1 of the MSH and what was observed in Experiment 3a, the $12 \pm 10\%$ higher mean percentage of trials with mediator shifts for the test group ($24 \pm 7\%$) than for the restudy group ($12 \pm 7\%$) was statistically significant, $F(1,46) = 5.982, p = .018, \eta^2_p = .115$. The main effect of phase was also significant, $F(2,46) = 19.299, p < .001, \eta^2_p = .296$, showing a decrease in mediator shifts across phases, as in Experiment 3a. Specifically, averaged over the test and restudy groups, the $13 \pm 8\%$ greater percentage of trials with mediator shifts on Review 1 ($28 \pm 8\%$) as compared to Review 2 ($15 \pm 4\%$) was significant ($p = <.001$), but the $3 \pm 3\%$ difference in the percentage of trials with
mediator shifts on Review 2 (15 ± 4%) as compared to Review 3 (12 ± 4%) was not significant 
($p = .158), F(2,46) = 1.951, p = .148, \eta^2 = .041$. The interaction between Review Phase and 
Type of Review was also not significant, $F(2,46) = 1.951, p = .148, \eta^2 = .041$.

The second section of the bottom half of Table 6 presents the mean percentage of trials 
with mediator shifts following target recall failure or success on the immediately preceding 
review phase. To assess Pyc and Rawson’s (2012) prediction that mediator shifts are more likely 
to occur after a previous recall failure (i.e., prediction 2 of the MSH), a 2 (Review Phase: Review 
2 vs. Review 3) x 2 (Previous Review Recall: Success vs. Failure) repeated-measures ANOVA 
was conducted on the percentage of trials with mediator shifts. Consistent with the MSH but 
unlike in Experiment 3a, mediator shifts were now more likely to occur following recall failure 
(13 ± 6%) than following recall success (5 ± 2%), $F(1,23) = 11.394, p = .003, \eta^2 = .331$. The 
interaction was also significant, demonstrating the significant 11 ± 7% greater percentage of 
mediator shifts following recall failure versus recall success in Review 2 was significantly 
different from the significant 4 ± 4% smaller percentage of mediator shifts following recall 
failure versus recall success in Review 3, $F(1,23) = 5.412, p = .029, \eta^2 = .190$. The greater 
percentage of mediator shifts in Review 2 following recall failure than following recall success 
stands in contrast to Experiment 3a, in which mediator shifts were equally likely to occur 
following a recall failure and recall success and is now congruent with the MSH.

**Final test recall.** The last column of the bottom half of Table 4 presents “specified” 
mediator recall for the final test. As mentioned above, “specified” mediator recall refers to 
participants recalling the mediators that were I created to be phonologically related to the cue 
and semantically related to the target, such as “ring” for bring-finger. Unlike in Experiment 3a, 
the test group recalled 15 ± 14% more “specified” mediators on the final test (47 ± 9%) as
compared to the restudy group (32 ± 11%), \( t(46) = 2.096, p = .042, d = .61 \). This supports the MEH.

The last section of Table 2 presents the mean Session 2 final test correct recall percentages of a target to its cue and of the mediator that was reported in Review Phase 3 for that same cue. A 2 (Type of Review: Test vs. Restudy) × 2 (Type of Final Test Recall: Mediator vs. Target) mixed-factor ANOVA was conducted, with the first factor being between-participants and the second factor, within-participants. Consistent with Experiment 3a, correct mediator recall was scored using the participant’s most recently generated mediator (i.e., the mediator from Review 3). As in Experiment 3a, overall mediator recall (61 ± 9%) was significantly higher than target recall (50 ± 7%), \( F(1,46) = 15.661, p < .001, \eta_p^2 = .254 \). Also, averaged across mediator and target recall, there was a robust 22 ± 15% TE, as indicated by the significant main effect of Type of Review, \( F(1,46) = 8.788, p = .005, \eta_p^2 = .160 \).

As in Experiment 2, which used the same pairs, the interaction between Type of Review and Type of Final Test Recall was also significant, \( F(1,46) = 31.701, p < .001, \eta_p^2 = .408 \), due to a larger TE for target recall than mediator recall. As expected, the 37 ± 14% TE for target recall was significant, \( F(1,46) = 27.301, p < .001, \eta_p^2 = .372 \). However, the 7 ± 17% TE for mediator recall was not significant, \( F(1,46) = .622, p = .435, \eta_p^2 = .013 \). Similar to Experiment 3a, the nonsignificant TE for mediator recall fails to support the MEH and prediction 3 of the MSH, and suggests that the mediator shifts by the test group in the review phase did not translate to higher mediator recall in Session 2 as compared to the restudy group.

---

8 The 2 (Type of Review: Test vs. Restudy) × 2 (Type of Final Test Recall: Mediator vs. Target) mixed-factor ANOVA was repeated to assess if participants were recalling mediators that were not their most recently generated mediators. Mediator recall remained similar when the participants’ Review 1 mediators were considered as correct mediator recall (59 ± 9%) and when Review 2 mediators were considered as correct mediator recall (60 ± 8%).
The last section of Table 3 presents mean percent target recall in Session 2, conditionalized on whether or not the participants recalled the corresponding mediator. A 2 (Type of Review: Test vs. Restudy) × 2 (Mediator Recall: Mediator recalled vs. Mediator not recalled) mixed-factor ANOVA was conducted, with the first factor being between-participants and the second, within-participants. Similar to the results of the previous experiments, the Type of Review Test main effect was significant, $F(1,43) = 33.765, p < .001, \eta^2_p = .440$, based on the $31 \pm 11\%$ TE. Additionally, target recall to the cue was $50 \pm 11\%$ overall greater when the mediator could be recalled ($66 \pm 8\%$) as compared to when it could not be recalled ($17 \pm 8\%$), $F(1,43) = 82.569, p < .001, \eta^2_p = .658$, which is congruent with mediation. Furthermore, unlike in the previous experiments, there was now evidence that testing enhanced mediation. The TE for target recall was $24 \pm 22\%$ larger when the mediator reported in Review Phase 3 was recalled as compared to when it was not recalled, $t(43) = 2.148, p = .037, d = .643$. In contrast to the previous experiments, the results provide support (a) for the MEH with the current English word pairs and explicit mediation instructions and (b) for the MSH as an explanation for why testing enhances mediation. This conclusion is also supported by the computed Bayes Factor, although the evidence is only very weak anecdotal evidence per Van Doorn et al. (2019). Specifically, the data are 2.484 times more likely under a model that includes the interaction between Type of Review and Mediator Recall than one that includes only the two main effects.

---

9 Three participants (one from the test group and two from the restudy group) were excluded from the analysis because they recalled all of the mediators during Session 2.
10 The same pattern of results emerged when considering mediators recalled from Review 1 and 2, instead of the most recently reported mediator. There was a significant $22 \pm 21\%$ difference in TEs for target recall when the mediator from Review 1 could be recalled in Session 2 as compared to when it could not, and a significant $23 \pm 21\%$ difference in TEs for target recall when the mediator from Review 2 could be recalled in Session 2 as compared to when it could not.
The results suggest that foreign language translation materials are not necessary for the MEH and MSH to apply, but rather, mediation must be explicitly instructed and the mediation must be based on phonological-semantic/associative mediators, which were employed in Pyc and Rawson (2010).

**Experiments 2 and 3b**

To assess the effects of explicit mediation instructions with phonologically related cues and mediators, a 2 (Experiment: 2 vs. 3b) × 2 (Type of Review: Test vs. Restudy) × 2 (Mediator Recall: Mediator recalled vs. not recalled) mixed-factor ANOVA was conducted on target recall with the first two factors being between-participants and the last factor, within-participants. When the data from Experiments 2 and 3b were combined, the overall 27 ± 8% TE was significant, with the overall 55 ± 5% target recall in the test group being greater than the 28 ± 5% recall in the restudy group, $F(1,160) = 50.719, p < .001, \eta_p^2 = .241$. Furthermore, target recall to the cue was overall 37 ± 5% higher when the mediator could be recalled (60 ± 5%) as compared to when it could not be recalled (23 ± 4%), $F(1,160) = 204.390, p < .001, \eta_p^2 = .561$, congruent with mediation. Consistent with the MEH, averaged across Experiments 2 and 3b, the TE for target recall was larger when the mediator could be recalled as compared to when it could not be recalled, as shown by the significant 15 ± 10% difference in TEs for target recall when the mediator could be recalled as compared to when it could not, $t(162) = 2.807, p = .006, d = .457$. This conclusion is numerically supported, though barely so, by the computed Bayes Factor, although the evidence is very weak and anecdotal (Van Doorn et al., 2019). Specifically, the data are only 1.377 times more likely under a model with the main effects of Type of Review and Mediator Recall and the interaction between Type of Review and Mediator Recall, as compared to a model with only the two main effects of Type of Review and Mediator Recall. There was
also a significant Experiment × Mediator Recall interaction showing that the difference in target recall when the mediator could be recalled versus when it could not be recalled was larger for Experiment 3b than for Experiment 2, \( F(1,160) = 23.953, p < .001, \eta^2_p = .130 \). This likely was due to the difference in instructions between the two experiments. The participants in Experiment 3b were explicitly told to use mediation, whereas any mediation that would occur in Experiment 2 would be spontaneous. The three-way interaction was marginally significant, \( F(1,160) = 3.053, p = .083, \eta^2_p = .019 \), suggesting that testing may enhance mediation more when participants are required to report mediators than when they are not. The Bayes Factor showed that the best model is a model with main effects of Experiment, Type of Review, and Mediator Recall, an interaction between Mediator Recall and Type of Review, and an interaction between Mediator Recall and Experiment. This model is 3.220 times more likely than a model containing the three-way interaction; however, this value only just passes the threshold for moderate evidence, per Van Doorn et al. (2019).

**Experiments 3a and 3b**

To compare semantic-semantic mediation and phonological-semantic mediation when participants are instructed to use explicit mediation instructions, a 2 (Experiment: 3a vs. 3b) × 2 (Type of Review: Test vs. Restudy) × 2 (Mediator Recall: Mediator recalled vs. not recalled) mixed-factor ANOVA was conducted on target recall with the first two factors being between-participants and the last factor, within-participants. When the data from Experiments 3a and 3b were combined, the overall 32 ± 7% TE was significant, with the overall 62 ± 7% target recall in the test group being greater than the 30 ± 5% recall in the restudy group, \( F(1,87) = 74.661, p < .001, \eta^2_p = .462 \). Target recall was also 10 ± 7% higher overall in Experiment 3a than in Experiment 3b, \( F(1,87) = 6.678, p = .011, \eta^2_p = .071 \). Furthermore, target recall to the cue was
overall 45 ± 7% higher when the mediator could be recalled (69 ± 5%) as compared to when it could not be recalled (24 ± 6%), $F(1,87) = 146.423, p < .001, \eta^2_p = .627$, congruent with mediation having been employed. Consistent with the MEH, averaged across Experiments 3a and 3b, there was a marginally significant 13 ± 15% difference in TEs for target recall when the mediator could be recalled (38 ± 10%) as compared to when it could not (25 ± 11%), $F(1,87) = 3.325, p = .072, \eta^2_p = .037$. The three-way interaction was not significant, however, $F(1,87) = 1.871, p = .175, \eta^2_p = .021$. Thus, even though the data supported the MEH with explicit mediation instructions for phonological-semantic mediation (Experiment 3b) but not for semantic-semantic mediation (Experiment 3a) when the data for the two different types of mediation were analyzed separately, the nonsignificant three-way interaction in the combined analysis suggests that one should be cautious in concluding that differential support for the MEH is provided by phonological-semantic and semantic-semantic mediation when participants are explicitly instructed to use mediational strategies.

**General Discussion**

Pyc and Rawson’s (2010) MEH was assessed across a series of three experiments. The strongest version of the MEH states that the TE for cue-target pairs is based on the following: Relative to restudy of a cue-target pair, a test on the cue for recall of its target (a) results in stronger activation of a mediator, (b) creates a stronger cue-to-mediator association, (c) creates a stronger mediator-to-target association and (d) thereby facilitates the later retrieval of a target to its cue via the mediational chain that is used during that later retrieval. A weaker version of the MEH would not include component (b). (The distinction between strong and weak versions of the MEH is mine; Pyc and Rawson did not make it nor did they explicitly delineate the four components.) Pyc and Rawson’s (2010, 2012) data supported the strongest version of the MEH
using Swahili-English word pairs and explicit instructions for participants to generate and report the mediator they used for each cue-target pair when they studied it or were tested on it. In the final test, in the group that was given the most rigorous evaluation of the MEH, participants were given the cue and asked to recall both their most recently generated mediator and the target.

Assuming a standard cue-target TE emerges, the strongest version of the MEH predicts (1) a significant TE for cue-mediator recall, (2) a significant TE for mediator-target recall, (3) better recall of the target when the mediator could be recalled as compared to when it could not be recalled (this demonstrates that mediation occurred) and (4) a larger TE for target recall when the mediator could be recalled than when it could not be recalled. Pyc and Rawson’s results confirmed all of these predictions.

The current experiments further examined the MEH with English word pairs having weak associative and semantic relatedness. Most important, I compared results for spontaneous mediation conditions (Experiments 1 and 2) and explicit mediation instructions (Experiments 3a and 3b). As expected, a standard TE for cue-target recall emerged in all three experiments. However, disconfirming Prediction 1 of the strongest version of the MEH, a significant TE for cue-mediator recall did not emerge in any of the current experiments. This suggests that testing did not strengthen the mediators as compared to restudying. However, as mentioned above, the weaker version of the MEH can still be supported without this component.

Prediction 2, which predicts a significant TE for mediator-target recall, would be directly tested by asking participants to recall the target to the mediator (i.e., the “specified” mediator for Experiments 1 and 2, and their most recently generated mediator for Experiment 3), which the current experiments did not do. However, the current experiments indirectly tested Prediction 2, by asking participants to recall both the mediator and the target to the cue. Experiments 1-3a
indirectly disconfirmed Prediction 2, because there was no difference in the TE for target recall when the mediator could be recalled as compared to when it could not be recalled, suggesting that testing did not strengthen the mediator-target link. In contrast, Experiment 3b demonstrated a larger TE for target recall when the mediator could be recalled as compared to when it could not be recalled, suggesting indirect evidence for Prediction 2 of the MEH.

Prediction 3 states that better recall of the target when the mediator could be recalled should emerge as compared to when the mediator could not be recalled. In all three experiments, target recall was higher overall when the mediator could be recalled as compared to when it could not be recalled. This confirms Prediction 3 and is congruent with mediation having occurred. However, more important is Prediction 4, which states that there will be a larger TE for target recall when the mediator could be recalled than when it could not be recalled. This prediction was only confirmed in Experiment 3b, which used explicit mediation instructions and English word pairs with mediators that are phonologically related to the cue and semantically/associatively related to the target (created to be similar to the word pairs used in Pyc and Rawson’s, 2010 study). Therefore, mediation occurred in all experiments, but only in Experiment 3b could mediation explain the cue-target TE.

Experiments 3a and 3b also allowed for the predictions of the MSH to be tested. The MSH complements the MEH by proposing that testing enhances the beneficial effects of mediation by prompting participants to shift to new, more effective mediators. The MSH is important because it provides a reason as to why testing may enhance the beneficial effects of mediation as compared to restudying. The MSH predicts: (1) For multiple study/review cycles, participants in the test group will shift mediators more often in the test review phases than participants in the restudy group; (2) In the test group, participants are more likely to shift
mediators after a previous retrieval failure as compared to a previous retrieval success; (3) Participants in the test group should have better mediator recall at final test as compared to participants in the restudy group. Experiments 1 and 2 did not allow for the MSH to be tested because participants did not report their mediators. The MSH was not supported in Experiment 3a, which is not surprising, given that the MEH was also not supported and that the MSH complements the MEH. Although the Experiment 3a finding that participants who received a review test in Session 1 did shift mediators more often than participants who received restudy review did favor the MSH (i.e, Prediction 1 was confirmed), contrary to the MSH, these shifts did not depend on previous recall failures nor did they lead to higher mediator recall in Session 2 (i.e., Predictions 2 and 3 were disconfirmed). In contrast, the Experiment 3b results supported Prediction 2 of the MSH as well as Prediction 1. The test group was more likely to shift mediators than the restudy group, and mediator shifts were more likely to occur after recall failures. This suggests that testing may have enhanced the beneficial effect of mediation in Experiment 3b (i.e., the MEH was supported) in part by prompting participants to engage in more mediation shifts. However, contrary to the MSH’s Prediction 3, the test group was not more likely to recall the mediator reported in the most recent Phase 3 review as compared to the restudy group.

The findings from the current experiments suggest that testing-enhanced mediation does not occur under spontaneous mediation conditions, even with multiple review cycles that provide the opportunity for TPL and mediator shifting to occur. This finding is consistent with Cho, Neely, Brennan, et al. (2017), who also failed to find evidence for the MEH under spontaneous mediation conditions. In contrast to the current experiments, Cho, Neely, Brennan, et al.’s experiment only had one study/review cycle with no feedback, providing no opportunity for TPL
and mediator shifting to occur. This suggests that adding the opportunity for TPL does not enhance testing-enhanced mediation strategies when the conditions are spontaneous. The lack of evidence for the MEH under spontaneous mediation conditions demonstrates the limitations of the MEH as a general explanation of the TE observed for paired associates and suggests that there are other underlying explanations for cue-target TEs that need to be explored. Future experiments could provide participants with a questionnaire at the end of the session regarding the strategies they had previously used in learning and retrieving the cue-target associations, to assess if strategies other than mediation were being used. For example, it is possible that participants were using a strategy that could link the cues and targets together without a mediator, such as imagining an image containing both the cue and target.

Experiment 3 suggests that testing-enhanced mediation can occur with explicit mediation instructions, but this depends on the materials. The MEH and MSH were not supported by the results of Experiment 3a, which used semantically and associatively weakly related English cue-target word pairs with a strong associate of the cue that could have been be used as a mediator. There could be at least two reasons for this. First, the weak semantic association between the cue and target may have been strong enough that mediation was simply not needed to learn the cue-target associations. In Experiment 3b, the word pairs were more difficult to learn than the cue-target pairs in Experiments 1 and 3a (as evidenced by lower review test accuracy), and mediation may have been useful for learning these cue-target word pairs.

Second, according to Pyc and Rawson (2012), providing participants with specific instructions on how to generate mediators can limit the number of effective mediators that participants can use, which could in turn limit the support for mediation. It is possible that if the instructions were not as limiting, participants may have switched mediators more frequently in
Experiment 3a in the review phases. For example, if participants experienced recall failure in Experiment 3a, they may not have been able to come up with a new effective mediator to switch to that still fit within the study instructions. Importantly, ten of the items in Experiment 3b had multiple mediators that fit the generation instructions, making it possible for participants to switch mediators more than in Experiment 3a. Future experiments could assess whether testing-enhanced mediation occurs and is greater when mediation instructions are less specific and limiting.

It is also not clear from the current experiments if testing-enhanced mediation would still occur under the same circumstances as Experiment 3b, but with no opportunity for TPL. Pyc and Rawson (2010, 2012) and the current Experiment 3b supported the MEH with explicit instructions and materials with a phonological link between the cues and mediators, but with multiple review cycles that allow for TPL. Therefore, mediation could be influencing both encoding and retrieval, which could enhance the overall mediation effect. Future studies could address this by replicating Experiment 3b, but with only one study/review cycle, and no corrective feedback on the test review cycles. It may be that the MEH is not supported under the conditions set in Experiment 3b without multiple study/review cycles.

All of the current experiments suggested that mediation in general was occurring, even though its deployment was not enhanced by review testing. This is consistent with previous research suggesting that mediational strategies in general can be beneficial for connecting difficult to learn concepts. Bugelski (1968) examined a mediational strategy with paired-associate learning and suggested that participants can learn word pairs better when using an image as a mediator. Atkinson (1974) later termed this the keyword method and demonstrated its effectiveness with Russian-English word pairs. Participants were required to come up with a
word that sounded like the Russian cue (i.e., the keyword) and then to create a mental image of the keyword’s referent interacting with the target’s referent (i.e., the English translation of the Russian cue). Participants who were required to use this imagery-based keyword method learned the pairs better than participants in a control group that did not receive mediation instructions.

The keyword method has also been evaluated in a classroom setting, typically when learning foreign language word pairs (e.g., Dolean, 2014; Wyra et al., 2007), although it has also been utilized in other learning situations, such as teaching children to connect states and their capitals (Levin et al., 1980), to learn science concepts (Rosenheck, Levin, & Levin, 1989), and to connect people’s names with accomplishments (Shriberg, Levin, McCormick, & Pressley, 1982). Given the effectiveness of the keyword method in classroom settings, future studies should assess whether a similar mediation strategy can be enhanced by review testing in the classroom.

It is well established that the TE can extend to a classroom setting, but to my knowledge, it is not yet clear what the exact nature of the mediational strategies that benefit learning and retrieval is and how testing enhances their deployment and/or beneficial effects. When students attempt to use a mediation strategy to link together difficult concepts, they could be encouraged to switch to a new mediation strategy if they experience recall failure during review. Additionally, it should be explored which materials and conditions (i.e., spontaneous or explicit) should be used in a classroom setting to produce the greatest benefit for learning and retrieval.

Overall, the current studies add to the literature by investigating the conditions and materials that will yield support for the MEH and MSH. The current studies show that (a) the MEH/MSH can be applied to materials other than foreign language translation equivalents, depending on how the materials are constructed and the instructions that are given, and (b) explicit mediation instructions are necessary for testing-enhanced mediation to occur. The
current studies also demonstrate boundary conditions for the MEH and show the need to develop other explanations of the TE in paired associate learning.
References


Cadaret, C. N., & Yates, D. T. (2018). Retrieval practice in the form of online homework improved information retention more when spaced 5 days rather than 1 day after class in two physiology courses. Advanced Physiological Education, 42, 305-310. doi: 10.1152/advan.00104.2017


JASP Team (2019). JASP (Version 0.11.1) [Computer software].


Odegard, T. N., & Koen, J. D. (2007). “None of the above” as a correct and incorrect alternative on a multiple-choice test: Implications for the testing effect. *Memory, 15*(8), 873-885. doi: 10.1080/09658210701746621


Appendix A

FAS, BAS, and LSA values for each triad in Experiment 1 and 3a.

<table>
<thead>
<tr>
<th>Cue</th>
<th>Mediator</th>
<th>Target</th>
<th>FAS</th>
<th>BAS</th>
<th>LSA</th>
<th>FAS</th>
<th>BAS</th>
<th>LSA</th>
<th>FAS</th>
<th>BAS</th>
<th>LSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>nurse</td>
<td>doctor</td>
<td>lawyer</td>
<td>0.54</td>
<td>0.38</td>
<td>0.52</td>
<td>0.10</td>
<td>0.15</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>eyes</td>
<td>nose</td>
<td>smell</td>
<td>0.16</td>
<td>0.00</td>
<td>0.59</td>
<td>0.10</td>
<td>0.12</td>
<td>0.58</td>
<td>0.00</td>
<td>0.00</td>
<td>0.42</td>
</tr>
<tr>
<td>couch</td>
<td>chair</td>
<td>sit</td>
<td>0.28</td>
<td>0.11</td>
<td>0.69</td>
<td>0.21</td>
<td>0.18</td>
<td>0.62</td>
<td>0.05</td>
<td>0.00</td>
<td>0.51</td>
</tr>
<tr>
<td>pants</td>
<td>shirt</td>
<td>collar</td>
<td>0.18</td>
<td>0.27</td>
<td>0.78</td>
<td>0.03</td>
<td>0.34</td>
<td>0.56</td>
<td>0.00</td>
<td>0.00</td>
<td>0.51</td>
</tr>
<tr>
<td>soap</td>
<td>water</td>
<td>drink</td>
<td>0.09</td>
<td>0.00</td>
<td>0.39</td>
<td>0.16</td>
<td>0.15</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.19</td>
</tr>
<tr>
<td>twinkle</td>
<td>star</td>
<td>wish</td>
<td>0.66</td>
<td>0.03</td>
<td>0.52</td>
<td>0.05</td>
<td>0.12</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.22</td>
</tr>
<tr>
<td>lemon</td>
<td>sour</td>
<td>sweet</td>
<td>0.16</td>
<td>0.09</td>
<td>0.70</td>
<td>0.40</td>
<td>0.37</td>
<td>0.45</td>
<td>0.00</td>
<td>0.00</td>
<td>0.35</td>
</tr>
<tr>
<td>wrist</td>
<td>watch</td>
<td>clock</td>
<td>0.34</td>
<td>0.02</td>
<td>0.23</td>
<td>0.12</td>
<td>0.09</td>
<td>0.26</td>
<td>0.00</td>
<td>0.00</td>
<td>0.11</td>
</tr>
<tr>
<td>birthday</td>
<td>cake</td>
<td>pie</td>
<td>0.19</td>
<td>0.09</td>
<td>0.50</td>
<td>0.07</td>
<td>0.17</td>
<td>0.62</td>
<td>0.00</td>
<td>0.00</td>
<td>0.27</td>
</tr>
<tr>
<td>valley</td>
<td>mountain</td>
<td>peak</td>
<td>0.19</td>
<td>0.02</td>
<td>0.57</td>
<td>0.02</td>
<td>0.25</td>
<td>0.71</td>
<td>0.00</td>
<td>0.00</td>
<td>0.42</td>
</tr>
<tr>
<td>fast</td>
<td>slow</td>
<td>turtle</td>
<td>0.59</td>
<td>0.53</td>
<td>0.70</td>
<td>0.11</td>
<td>0.37</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.21</td>
</tr>
<tr>
<td>cry</td>
<td>baby</td>
<td>bottle</td>
<td>0.10</td>
<td>0.12</td>
<td>0.52</td>
<td>0.04</td>
<td>0.03</td>
<td>0.20</td>
<td>0.00</td>
<td>0.00</td>
<td>0.24</td>
</tr>
<tr>
<td>oyster</td>
<td>pearl</td>
<td>necklace</td>
<td>0.05</td>
<td>0.13</td>
<td>0.20</td>
<td>0.27</td>
<td>0.09</td>
<td>0.18</td>
<td>0.00</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>winter</td>
<td>snow</td>
<td>0.39</td>
<td>0.37</td>
<td>0.86</td>
<td>0.20</td>
<td>0.03</td>
<td>0.67</td>
<td>0.00</td>
<td>0.00</td>
<td>0.43</td>
</tr>
<tr>
<td>blade</td>
<td>razor</td>
<td>shave</td>
<td>0.23</td>
<td>0.29</td>
<td>0.52</td>
<td>0.17</td>
<td>0.15</td>
<td>0.30</td>
<td>0.02</td>
<td>0.00</td>
<td>0.17</td>
</tr>
<tr>
<td>minute</td>
<td>hour</td>
<td>glass</td>
<td>0.26</td>
<td>0.47</td>
<td>0.51</td>
<td>0.02</td>
<td>0.00</td>
<td>0.11</td>
<td>0.00</td>
<td>0.00</td>
<td>0.17</td>
</tr>
<tr>
<td>bull</td>
<td>cow</td>
<td>milk</td>
<td>0.55</td>
<td>0.10</td>
<td>0.21</td>
<td>0.35</td>
<td>0.39</td>
<td>0.60</td>
<td>0.00</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>window</td>
<td>door</td>
<td>knob</td>
<td>0.14</td>
<td>0.16</td>
<td>0.66</td>
<td>0.13</td>
<td>0.67</td>
<td>0.37</td>
<td>0.00</td>
<td>0.00</td>
<td>0.28</td>
</tr>
<tr>
<td>day</td>
<td>night</td>
<td>dark</td>
<td>0.81</td>
<td>0.69</td>
<td>0.54</td>
<td>0.19</td>
<td>0.21</td>
<td>0.59</td>
<td>0.00</td>
<td>0.00</td>
<td>0.28</td>
</tr>
<tr>
<td>wedding</td>
<td>ring</td>
<td>finger</td>
<td>0.04</td>
<td>0.10</td>
<td>0.22</td>
<td>0.20</td>
<td>0.06</td>
<td>0.36</td>
<td>0.00</td>
<td>0.00</td>
<td>0.19</td>
</tr>
<tr>
<td>school</td>
<td>bus</td>
<td>stop</td>
<td>0.07</td>
<td>0.08</td>
<td>0.32</td>
<td>0.06</td>
<td>0.00</td>
<td>0.47</td>
<td>0.00</td>
<td>0.00</td>
<td>0.20</td>
</tr>
<tr>
<td>beer</td>
<td>wine</td>
<td>grape</td>
<td>0.08</td>
<td>0.19</td>
<td>0.85</td>
<td>0.06</td>
<td>0.10</td>
<td>0.17</td>
<td>0.00</td>
<td>0.00</td>
<td>0.16</td>
</tr>
<tr>
<td>sport</td>
<td>baseball</td>
<td>glove</td>
<td>0.05</td>
<td>0.00</td>
<td>0.70</td>
<td>0.02</td>
<td>0.06</td>
<td>0.66</td>
<td>0.00</td>
<td>0.00</td>
<td>0.39</td>
</tr>
<tr>
<td>moon</td>
<td>sun</td>
<td>hot</td>
<td>0.29</td>
<td>0.15</td>
<td>0.28</td>
<td>0.08</td>
<td>0.00</td>
<td>0.34</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
</tr>
</tbody>
</table>

|        |          |        | 0.27| 0.18| 0.52| 0.13| 0.17| 0.39| 0.00| 0.00| 0.25|
Appendix B

FAS, BAS, LSA, and ESGQ values for each triad in Experiments 2 and 3b.

<table>
<thead>
<tr>
<th>Cue</th>
<th>Mediator</th>
<th>Target</th>
<th>FAS</th>
<th>BAS</th>
<th>LSA</th>
<th>FAS</th>
<th>BAS</th>
<th>LSA</th>
<th>ESGQ</th>
<th>FAS</th>
<th>BAS</th>
<th>LSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>toes</td>
<td>nose</td>
<td>smell</td>
<td>0.00</td>
<td>0.00</td>
<td>0.42</td>
<td>0.11</td>
<td>0.12</td>
<td>0.58</td>
<td>0.19</td>
<td>0.01</td>
<td>0.00</td>
<td>0.19</td>
</tr>
<tr>
<td>dare</td>
<td>chair</td>
<td>sit</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
<td>0.21</td>
<td>0.18</td>
<td>0.62</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.36</td>
</tr>
<tr>
<td>off</td>
<td>cough</td>
<td>choke</td>
<td>0.00</td>
<td>0.00</td>
<td>0.13</td>
<td>0.12</td>
<td>0.20</td>
<td>0.13</td>
<td>0.41</td>
<td>0.00</td>
<td>0.00</td>
<td>0.27</td>
</tr>
<tr>
<td>dirt</td>
<td>shirt</td>
<td>collar</td>
<td>0.00</td>
<td>0.00</td>
<td>0.28</td>
<td>0.03</td>
<td>0.34</td>
<td>0.56</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.14</td>
</tr>
<tr>
<td>glow</td>
<td>blow</td>
<td>bubbles</td>
<td>0.00</td>
<td>0.00</td>
<td>0.19</td>
<td>0.01</td>
<td>0.03</td>
<td>0.21</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.23</td>
</tr>
<tr>
<td>far</td>
<td>star</td>
<td>wish</td>
<td>0.00</td>
<td>0.00</td>
<td>0.22</td>
<td>0.05</td>
<td>0.13</td>
<td>0.06</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.28</td>
</tr>
<tr>
<td>flour</td>
<td>sour</td>
<td>sweet</td>
<td>0.00</td>
<td>0.00</td>
<td>0.36</td>
<td>0.41</td>
<td>0.37</td>
<td>0.45</td>
<td>0.17</td>
<td>0.00</td>
<td>0.00</td>
<td>0.47</td>
</tr>
<tr>
<td>blush</td>
<td>brush</td>
<td>hair</td>
<td>0.00</td>
<td>0.00</td>
<td>0.08</td>
<td>0.44</td>
<td>0.21</td>
<td>0.31</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.23</td>
</tr>
<tr>
<td>ache</td>
<td>cake</td>
<td>pie</td>
<td>0.00</td>
<td>0.00</td>
<td>0.11</td>
<td>0.07</td>
<td>0.17</td>
<td>0.62</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.13</td>
</tr>
<tr>
<td>dove</td>
<td>love</td>
<td>kiss</td>
<td>0.00</td>
<td>0.00</td>
<td>0.24</td>
<td>0.06</td>
<td>0.23</td>
<td>0.48</td>
<td>0.38</td>
<td>0.00</td>
<td>0.00</td>
<td>0.24</td>
</tr>
<tr>
<td>loft</td>
<td>soft</td>
<td>cotton</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>0.02</td>
<td>0.17</td>
<td>0.16</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.20</td>
</tr>
<tr>
<td>maybe</td>
<td>baby</td>
<td>bottle</td>
<td>0.00</td>
<td>0.00</td>
<td>0.23</td>
<td>0.05</td>
<td>0.02</td>
<td>0.20</td>
<td>0.62</td>
<td>0.00</td>
<td>0.00</td>
<td>0.26</td>
</tr>
<tr>
<td>girl</td>
<td>pearl</td>
<td>necklace</td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
<td>0.27</td>
<td>0.09</td>
<td>0.18</td>
<td>0.49</td>
<td>0.00</td>
<td>0.00</td>
<td>0.24</td>
</tr>
<tr>
<td>hall</td>
<td>ball</td>
<td>bounce</td>
<td>0.00</td>
<td>0.00</td>
<td>0.23</td>
<td>0.06</td>
<td>0.56</td>
<td>0.64</td>
<td>0.24</td>
<td>0.00</td>
<td>0.00</td>
<td>0.22</td>
</tr>
<tr>
<td>search</td>
<td>church</td>
<td>bell</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.03</td>
<td>0.05</td>
<td>0.18</td>
<td>0.12</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>now</td>
<td>cow</td>
<td>milk</td>
<td>0.00</td>
<td>0.00</td>
<td>0.23</td>
<td>0.35</td>
<td>0.39</td>
<td>0.60</td>
<td>0.12</td>
<td>0.00</td>
<td>0.00</td>
<td>0.13</td>
</tr>
<tr>
<td>more</td>
<td>floor</td>
<td>carpet</td>
<td>0.00</td>
<td>0.00</td>
<td>0.19</td>
<td>0.07</td>
<td>0.16</td>
<td>0.57</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.21</td>
</tr>
<tr>
<td>site</td>
<td>night</td>
<td>dark</td>
<td>0.00</td>
<td>0.00</td>
<td>0.10</td>
<td>0.19</td>
<td>0.21</td>
<td>0.59</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>bring</td>
<td>ring</td>
<td>finger</td>
<td>0.00</td>
<td>0.00</td>
<td>0.26</td>
<td>0.20</td>
<td>0.07</td>
<td>0.36</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.24</td>
</tr>
<tr>
<td>fly</td>
<td>lie</td>
<td>cheat</td>
<td>0.00</td>
<td>0.00</td>
<td>0.23</td>
<td>0.12</td>
<td>0.18</td>
<td>0.15</td>
<td>0.11</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>fuss</td>
<td>bus</td>
<td>stop</td>
<td>0.00</td>
<td>0.00</td>
<td>0.20</td>
<td>0.06</td>
<td>0.00</td>
<td>0.47</td>
<td>0.40</td>
<td>0.00</td>
<td>0.00</td>
<td>0.28</td>
</tr>
<tr>
<td>mine</td>
<td>wine</td>
<td>grape</td>
<td>0.00</td>
<td>0.00</td>
<td>0.13</td>
<td>0.06</td>
<td>0.10</td>
<td>0.17</td>
<td>0.09</td>
<td>0.00</td>
<td>0.00</td>
<td>0.13</td>
</tr>
<tr>
<td>one</td>
<td>sun</td>
<td>hot</td>
<td>0.00</td>
<td>0.00</td>
<td>0.24</td>
<td>0.09</td>
<td>0.00</td>
<td>0.34</td>
<td>0.11</td>
<td>0.00</td>
<td>0.00</td>
<td>0.25</td>
</tr>
<tr>
<td>fountain</td>
<td>mountain</td>
<td>peak</td>
<td>0.00</td>
<td>0.00</td>
<td>0.09</td>
<td>0.02</td>
<td>0.25</td>
<td>0.71</td>
<td>0.16</td>
<td>0.00</td>
<td>0.00</td>
<td>0.42</td>
</tr>
</tbody>
</table>

|          |           |        | 0.00| 0.00| 0.21| 0.13| 0.18| 0.39| 0.16 | 0.00| 0.00| 0.22 |
Table 1. Mean Session 1 correct recall percentages for the three review phases.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Review 1</th>
<th>Review 2</th>
<th>Review 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>52 ± 5%</td>
<td>83 ± 4%</td>
<td>93 ± 2%</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>29 ± 4%</td>
<td>52 ± 6%</td>
<td>68 ± 6%</td>
</tr>
<tr>
<td>Experiment 3a</td>
<td>72 ± 8%</td>
<td>87 ± 3%</td>
<td>94 ± 3%</td>
</tr>
<tr>
<td>Experiment 3b</td>
<td>42 ± 9%</td>
<td>61 ± 9%</td>
<td>71 ± 9%</td>
</tr>
</tbody>
</table>
Table 2. Mean Session 2 final test correct recall percentages of mediators and targets to their corresponding cues.

<table>
<thead>
<tr>
<th></th>
<th>Mediator</th>
<th>Target</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>87%</td>
<td>79%</td>
<td>7 ± 6%</td>
</tr>
<tr>
<td>Restudy</td>
<td>86%</td>
<td>44%</td>
<td>42 ± 6%</td>
</tr>
<tr>
<td>TE</td>
<td>1 ± 4%</td>
<td>35 ± 9%</td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>32%</td>
<td>47%</td>
<td>-15 ± 6%</td>
</tr>
<tr>
<td>Restudy</td>
<td>35%</td>
<td>26%</td>
<td>8 ± 6%</td>
</tr>
<tr>
<td>TE</td>
<td>-2 ± 8%</td>
<td>21 ± 8%</td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 3a</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>76%</td>
<td>82%</td>
<td>6 ± 5%</td>
</tr>
<tr>
<td>Restudy</td>
<td>71%</td>
<td>45%</td>
<td>26 ± 5%</td>
</tr>
<tr>
<td>TE</td>
<td>5 ± 11%</td>
<td>37 ± 10%</td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 3b</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>64%</td>
<td>69%</td>
<td>5 ± 8%</td>
</tr>
<tr>
<td>Restudy</td>
<td>58%</td>
<td>32%</td>
<td>-26 ± 8%</td>
</tr>
<tr>
<td>TE</td>
<td>7 ± 17%</td>
<td>37 ± 14%</td>
<td></td>
</tr>
</tbody>
</table>

Note: In Experiments 3a and 3b, correct mediator recall is defined as recall of the word that was reported in Review Phase 3 as being the mediator for the to-be-learned cue-target pair.
Table 3. Mean Session 2 recall percentages of a target to its cue, conditionalized on recall of the mediator (M).

<table>
<thead>
<tr>
<th></th>
<th>Experiment 1</th>
<th></th>
<th>Experiment 2</th>
<th></th>
<th>Experiment 3a</th>
<th></th>
<th>Experiment 3b</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M Recalled</td>
<td>M Not Recalled</td>
<td>Difference</td>
<td>M Recalled</td>
<td>M Not Recalled</td>
<td>Difference</td>
<td>M Recalled</td>
<td>M Not Recalled</td>
</tr>
<tr>
<td>Test</td>
<td>79%</td>
<td>69%</td>
<td>10 ± 8%</td>
<td>67%</td>
<td>40%</td>
<td>27 ± 7%</td>
<td>87%</td>
<td>46%</td>
</tr>
<tr>
<td>Restudy</td>
<td>44%</td>
<td>39%</td>
<td>5 ± 8%</td>
<td>42%</td>
<td>20%</td>
<td>22 ± 7%</td>
<td>54%</td>
<td>16%</td>
</tr>
<tr>
<td>TE</td>
<td>35 ± 9%</td>
<td>30 ± 14%</td>
<td></td>
<td>25 ± 11%</td>
<td>20 ± 8%</td>
<td></td>
<td>33 ± 13%</td>
<td>30 ± 16%</td>
</tr>
</tbody>
</table>

Note: In Experiments 3a and 3b, the mediator is defined as the word that was reported in Review Phase 3 as being the mediator for the to-be-learned cue-target pair.
**Table 4.** Mean percentages of times the “specified” mediator was reported.

<table>
<thead>
<tr>
<th></th>
<th>Experiment 3a</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Study Phase</td>
<td>Review 1 Phase</td>
<td>Review 2 Phase</td>
<td>Review 3 Phase</td>
<td>Final Test</td>
</tr>
<tr>
<td>Test</td>
<td>47 ± 8%</td>
<td>47 ± 9%</td>
<td>48 ± 10%</td>
<td>47 ± 10%</td>
<td>46 ± 9%</td>
</tr>
<tr>
<td>Restudy</td>
<td>49 ± 7%</td>
<td>48 ± 7%</td>
<td>49 ± 7%</td>
<td>49 ± 7%</td>
<td>41 ± 6%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Experiment 3b</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Study Phase</td>
<td>Review 1 Phase</td>
<td>Review 2 Phase</td>
<td>Review 3 Phase</td>
<td>Final Test</td>
</tr>
<tr>
<td>Test</td>
<td>48 ± 12%</td>
<td>49 ± 11%</td>
<td>49 ± 10%</td>
<td>49 ± 11%</td>
<td>47 ± 9%</td>
</tr>
<tr>
<td>Restudy</td>
<td>39 ± 14%</td>
<td>40 ± 13%</td>
<td>41 ± 13%</td>
<td>42 ± 13%</td>
<td>32 ± 11%</td>
</tr>
</tbody>
</table>

**Note.** A “specified” mediator refers to a mediator that was specified in the study materials. For experiment 3a, this refers to a reported mediator that was a strong associate of its cue and associatively related to its target (e.g., “tiger” as the mediator for lion-stripes). For Experiment 3b, this refers to a reported mediator that was phonologically related to its cue and semantically related to its target (e.g., “cow” for now-milk).
Table 5. Total percentages of trials with mediator shifts, broken down by whether the shift was from a “specified mediator” or from an “unspecified” mediator.

<table>
<thead>
<tr>
<th></th>
<th>Review 1</th>
<th>Review 2</th>
<th>Review 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shift from a “Specified” Mediator</td>
<td>Shift from an “Unspecified” Mediator</td>
<td>Shift from a “Specified” Mediator</td>
</tr>
<tr>
<td>Test</td>
<td>3%</td>
<td>16%</td>
<td>1%</td>
</tr>
<tr>
<td>Restudy</td>
<td>2%</td>
<td>10%</td>
<td>1%</td>
</tr>
<tr>
<td>TE</td>
<td>1 ± 2%</td>
<td>6 ± 7%</td>
<td>0 ± 1%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Review 1</th>
<th>Review 2</th>
<th>Review 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shift from a “Specified” Mediator</td>
<td>Shift from an “Unspecified” Mediator</td>
<td>Shift from a “Specified” Mediator</td>
</tr>
<tr>
<td>Test</td>
<td>5%</td>
<td>32%</td>
<td>2%</td>
</tr>
<tr>
<td>Restudy</td>
<td>1%</td>
<td>18%</td>
<td>1%</td>
</tr>
<tr>
<td>TE</td>
<td>4 ± 4%</td>
<td>14 ± 16%</td>
<td>2 ± 2%</td>
</tr>
</tbody>
</table>

Note. A “specified” mediator refers to a mediator that was specified in the study materials. For experiment 3a, this refers to a reported mediator that was a strong associate of its cue and associatively related to its target (e.g., “tiger” as the mediator for lion-stripes). For Experiment 3b, this refers to a reported mediator that was phonologically related to its cue and semantically related to its target (e.g., “cow” for now-milk).

1The shift could have been to a “specified” mediator or to an “unspecified” mediator. The first number in the parentheses indicates the percentage of trials with mediator shifts from an “unspecified” mediator to another “unspecified” mediator, and the second number indicates the percentage of trials with mediator shifts from an “unspecified” mediator to the “specified” mediator.
Table 6. Total percentages of trials with mediator shifts, broken down by the test versus restudy groups, and whether or not the target was successfully recalled to its cue in the immediately preceding phase.

<table>
<thead>
<tr>
<th>Experiment 3a</th>
<th>Review 1</th>
<th>Review 2</th>
<th>Review 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>19%</td>
<td>13%</td>
<td>11%</td>
</tr>
<tr>
<td>Restudy</td>
<td>11%</td>
<td>6%</td>
<td>5%</td>
</tr>
<tr>
<td>TE</td>
<td>7 ± 8%</td>
<td>7 ± 7%</td>
<td>6 ± 5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Review 1</th>
<th>Review 2</th>
<th>Review 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous Retrieval Failure</td>
<td>---</td>
<td>7%</td>
</tr>
<tr>
<td>Previous Retrieval Success</td>
<td>---</td>
<td>6%</td>
</tr>
<tr>
<td>Difference</td>
<td>---</td>
<td>1 ± 5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment 3b</th>
<th>Review 1</th>
<th>Review 2</th>
<th>Review 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>37%</td>
<td>19%</td>
<td>16%</td>
</tr>
<tr>
<td>Restudy</td>
<td>19%</td>
<td>10%</td>
<td>7%</td>
</tr>
<tr>
<td>TE</td>
<td>18 ± 16%</td>
<td>9 ± 9%</td>
<td>9 ± 9%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Review 1</th>
<th>Review 2</th>
<th>Review 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous Retrieval Failure</td>
<td>---</td>
<td>15%</td>
</tr>
<tr>
<td>Previous Retrieval Success</td>
<td>---</td>
<td>4%</td>
</tr>
<tr>
<td>Difference</td>
<td>---</td>
<td>11 ± 7%</td>
</tr>
</tbody>
</table>
**Figure 1.** Procedures for Experiments 1 and 2.

<table>
<thead>
<tr>
<th>Group</th>
<th>Study Phase 1 (24 pairs, 2.5 s each)</th>
<th>Review Phase 1 (7 s each)</th>
<th>Study Phase 2 (24 pairs, 2.5 s each)</th>
<th>Review Phase 2 (7 s each)</th>
<th>Study Phase 3 (24 pairs, 2.5 s each)</th>
<th>Review Phase 3 (7 s each)</th>
<th>Cue-Mediator-Target Final Test (self-paced)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>Study Phase 1 (24 pairs, 2.5 s each)</td>
<td>Review Phase 1 (7 s each)</td>
<td>Study Phase 2 (24 pairs, 2.5 s each)</td>
<td>Review Phase 2 (7 s each)</td>
<td>Study Phase 3 (24 pairs, 2.5 s each)</td>
<td>Review Phase 3 (7 s each)</td>
<td>Cue-Mediator-Target Final Test (self-paced)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Restudy (cue-target)</td>
</tr>
</tbody>
</table>

48 hrs
**Figure 2.** Procedures for Experiment 3.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>Test (cue-_____) + cue-target + generate mediator</td>
<td>Test (cue-_____) + cue-target + generate mediator</td>
<td>Test (cue-_____) + cue-target + generate mediator</td>
<td>48 hrs</td>
<td>Test (cue-_____)</td>
</tr>
<tr>
<td>Restudy</td>
<td>Restudy (cue-target) + generate mediator</td>
<td>Restudy (cue-target) + generate mediator</td>
<td>Restudy (cue-target) + generate mediator</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>