Failures to Replicate Hyper-Retrieval-Induced Forgetting in Arithmetic Memory

A Thesis Submitted to the College of

Graduate Studies and Research

in Partial Fulfilment of the Requirements

for the Degree of Master of Arts

in the Department of Psychology

University of Saskatchewan

Saskatoon

By

Anna Maslany

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements of a postgraduate degree from the University of Saskatchewan, I agree that the libraries at this university may make it freely available for inspection. I further agree that permission for copying this thesis in any manner, in whole or in part for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in the absence, by the Head of the Department of the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Psychology

Arts Building

University of Saskatchewan

9 Campus Drive

Saskatoon, Saskatchewan

S7K 5A5

ABSTRACT

Campbell and Phenix (2009) observed retrieval-induced forgetting (RIF) (slower response time) for simple addition facts (e.g., 3 + 4) immediately following 40 retrieval-practice blocks of their multiplication counterparts (3 × 4 = ?). A subsequent single retrieval of the previously unpracticed multiplication problems, however, produced an RIF effect about twice as large for their addition counterparts. Thus, a single retrieval of a multiplication fact appeared to produce much larger RIF of the addition counterpart than did many multiplication retrieval-practice trials. In subsequent similar studies, however, this *hyper-RIF* effect was not observed (e.g., Campbell & Thompson, 2012). The current studies further investigated hyper-RIF in arithmetic. In Chapter 2 (Experiment 1), composition of operands (unique vs. common) and amount of multiplication practice (6 vs. 20 repetitions of each problem) were manipulated. Participants solved multiplication problems (4 × 7 = ?) and then were tested on their memory for the addition counterparts (4 + 7 = ?) and control additions. Chapter 3 (Experiment 2) attempted an exact replication of Campbell and Phenix. In both studies, hyper-RIF was not observed. The results confirm the basic RIF effect of multiplication retrieval practice on addition counterparts, but cast doubt on the on the reality of the hyper-RIF effect observed by Campbell and Phenix. It is concluded that the hyper-RIF effect reported by Campbell and Phenix is an elusive or non-existent phenomenon; consequently, it cannot at this time be considered an important result in the RIF literature.

ACKNOWLEDGEMENTS

First of all, I wish to thank my supervisor, Dr. Jamie Campbell for the great amount of time and effort he put forward to help me get the most out of this program. I am forever grateful. I would also like to thank Dr. Tammy Marche and Dr. Valerie Thompson for providing ideas and guidance.

I would like to thank my fiancé Dr. Jonathan Misskey for helping me through my struggles and talking on the phone with me every night to make me feel better. Also, thank you very much to my dad, Dr. George Maslany who provided me with endless amounts of advice and guidance, and to my mom, Sharon Wheale for helping me through my hard times and deciding what steps to take next. Also, thank you to my brother Dr. Jurgen Maslany who taught me how to continually work towards my goals. I have the most wonderful family and I am very grateful to them all for helping me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICIES

LIST OF ABREVIATIONS

MP	multiplication practiced
ms	milliseconds
MU	multiplication unpracticed
Nrp	unpracticed, unrelated items
RIF	retrieval-induced forgetting
Rp+	practiced items
Rp-	related, unpracticed items
RT	response time

CHAPTER 1

Failures to Replicate Hyper-Retrieval-Induced Forgetting in Arithmetic Memory

Remembering can cause forgetting (Anderson, Bjork & Bjork, 1994). Thsi counterintuitive phenomenon is called retrieval-induced forgetting (RIF) and has been a popular research topic over the past 20 years. Anderson et al., (1994) were the first researchers to empirically demonstrate RIF and this influential paper has been cited over 750 times. According to Anderson (2003), the typical amount of RIF that is observed in these studies ranges from 9% to 20%. A real world example of RIF occurs when one has a list of groceries in mind to buy from the corner store, most items are remembered, but some are forgotten. It is not until one gets home to put away the groceries that he or she realizes they have forgotten some of the items on the list. This is an everyday example of RIF because the very act of remembering some items caused the forgetting of others.

This document is composed of two experiments that investigated RIF in arithmetic facts. Specifically, these experiments were conducted to examine why a very high level of RIF occurred in one study (Campbell & Phenix, 2009), but not others (Campbell & Thompson, 2012). Experiment 1 combined the variables that differ across studies (i.e., Campbell & Phenix; Campbell & Thompson) to examine if different variable manipulations caused this large amount of RIF to occur. Experiment 1 is presented in detail and the results of this experiment set the stage for Experiment 2 and the ultimate conclusions of this thesis. Before these experiments are discussed in detail it is important to provide a historical review of RIF and other relevant studies that have empirically demonstrated the effect.

Historical Review of RIF

The study of RIF began with retrieval inhibition (i.e., when a memory cannot be retrieved because it is inhibited by actions such as encoding or retrieval). Retrieval inhibition can be demonstrated in experiments that have implemented the retroactive interference paradigm (Bjork, 1989). This paradigm is composed of several phases. In the first phase, participants learn a list of paired associate words (e.g., Dog—Rock), followed by a similar list with the same first word, but a different second word in the pair (e.g., Dog—Lamp). In the next phase of the experiment, participants are tested on their recall of both words that came second in the associate pairs (e.g., Rock; Lamp). Typical results suggest that recall of the word that was studied first is impaired compared to the word that was studied second, presumably because of the intervening study of the second word. Interestingly, this impairment does not occur to the same degree when the two paired associates do not share a common word (e.g., first, Road-Rock, second, Dog-Lamp). This paradigm has provided evidence that new information in memory overrides items that were recalled earlier, and this effect is larger for related memory items than unrelated items (Bjork, 1989).

Similar evidence that information is overwritten in memory can be found with the part-set cueing paradigm (Bjork, 1989). In the first phase of this paradigm, participants study a list of category names, and then are tested on their memory for these items in a category-cued recall test. Results demonstrate that the to-be-remembered items are recalled in the category-cued recall test (e.g., the cue Fruit facilitates recall of Apple, Orange, and Banana etc.). However, remembering these items often impairs memory for the items that were not cued, compared to the baseline items that have no cue associated with them. The effects of part-set cueing have been explained in different ways. Some researchers argue that the inability to recall previously studied items occurs because cues are strengthened at test, which blocks access to the other items that are not related to that cue (Rundus, 1973). Alternatively, others argue that the inability to recall items is caused by the reorganization of the cued items at test compared to how they were organized at study. This reorganization causes the impairment of related cued items at test (i.e., separating cue and non-cue words during study) (e.g., Basden & Basden, 1995). The cues create competition for related items, much like an inhibitory explanation of RIF (Anderson et al., 1994), to be discussed later.

RIF research has been investigated in a variety of different types of memory including episodic (e.g., Cirrani & Shimamura, 1999), recognition (e.g., Hicks & Starns, 2004; Spitzer & Bäuml, 2007) and semantic memory (e.g., Blaxton & Neely, 1983). It has been studied in many different contexts including autobiographical memories (e.g., Barnier, Hung & Conway, 2004), social cognition (e.g., Storm, Bjork & Bjork, 2005), eyewitness memory (e.g., Shaw, Bjork, & Handal, 1995), memory for location (e.g., Gómez-Ariza, Fernandez, & Bajo, 2012), and varied delays between study and test (Abel & Baüml, 2012). There are over 200 studies that have experimentally demonstrated RIF, but it is still poorly understood (Storm & Levy, 2012).

RIF is a phenomenon that has greatly influenced memory research in recent years. In RIF, repeated retrieval practice of a memory item can impair retrieval of related, unpracticed memory items. Here, memory is enhanced for practiced items; unpracticed, unrelated items are used as a baseline; and unpracticed, related memory items are most difficult to retrieve (i.e., RIF) (Anderson et al., 1994).

The majority of studies have investigated RIF using the retrieval practice paradigm (e.g., Anderson, et al., 1994; Anderson, Bjork & Bjork, 2000). In the first phase, participants are asked to study a list of category-exemplar pairs (e.g., Fruit-- Apple) and then practice retrieval on a subset of those items (e.g., Fruit-- Apple). The items that participants practice retrieval on are called Rp+ items. The unpracticed items provide a baseline (e.g., Animal -- Dog). These items do not share a category or an exemplar with those that were practiced and are referred to as Nrp items. After a delay, participants are given a final recall test using cued category-exemplar pairs (e.g., Fruit- Ap_____), and participants memory is measured by the number of correctly recalled items. The items that share a category with those that were practiced, but were not practiced themselves in the previous practice phase (i.e., related, unpracticed items) (e.g., Fruit- Banana) are referred to as the RP- items. RIF occurs when memory for RP- items is poor relative to the Nrp items.

Theoretical Explanations of RIF

RIF has been demonstrated in numerous memory tasks (Anderson, 2003; Anderson et al., 1994; see Storm & Levy, 2012 for a recent review). One theoretical explanation for RIF is that related memory items are inhibited if they are strong retrieval competitors to the target memory (e.g., Shivde & Anderson, 2001) (i.e., inhibition); alternatively, access to related items could encounter interference because of strengthening of some items from target practice, but not be actively inhibited (e.g., Camp, Pecher, & Schmidt, 2007) (i.e., interference). The interference theory of RIF suggests that forgetting occurs because of the introduction of new information into memory and new information competes with older information for memory access (e.g., Butler, William, Zachs, & Maki, 2001). According to the inhibitory account (e.g., Anderson et al., 1994; Anderson, 2003) forgetting occurs because of the competition between related items in memory and memory items are suppressed and forgotten (Bjork, 1989).

RIF in Semantic Memory

RIF has primarily been investigated using episodic tasks (i.e., memory for events) (e.g., Cirrani & Shimamura, 1999), but also with semantic tasks. The experiments conducted and reported in this document examine RIF of common, memorized arithmetic facts, which is semantic memory (e.g., 3 + 4) and therefore, it is important to review RIF in semantic memory.

One of the first studies to examine inhibition in semantic memory was conducted by Blaxton and Neely (1983). In this study, participants were presented with categories (e.g., Fish). Participants either generated the targets (e.g., Fish- B____ for the exemplar Bass) or studied them (e.g., Fish-Bass). The speed at which they could generate an exemplar associated with that category was measured (e.g., retrieve the exemplar Bass when presented with the category Fish). Prior to this phase, participants retrieved one or four exemplars from the same (e.g., Fish) or a different category (e.g., Fruit). Results suggested that when participants studied the category and exemplars, participants were faster when the exemplar was from the same category as the target, compared to when it was not (i.e., facilitation). When participants generated primes, facilitation occurred when they retrieved of one exemplar from the same category. In contrast, when participants generated items and retrieved four items from the same category, rather than one item, facilitation did not occur. Blaxton and Neely argued that like episodic memory, inhibition occurred when more than one exemplar and category was recalled from semantic memory.

Johnson and Anderson (2004) also examined inhibition in semantic memory. In two experiments, participants were presented with homographs to investigate semantic memory for these items (i.e., words that are spelt the same way, but have different meanings; e.g., wound means an injury and wound means to turn). Participants were asked to generate a meaning associated with a homograph. Participants retrieved one or eight items and were either given a one letter cue (Experiment 1) or two letter cue (Experiment 2). The results of Experiment 1 demonstrated that retrieval practice of one item caused that item to be more accessible at retrieval, but when more items were practiced, suppression effects occurred at retrieval for these items, compared to performance on baseline items. Experiment 2 found similar results, but performance on related items (i.e., practicing four or eight items during retrieval practice) caused suppression below baseline. Participants recalled fewer items after practicing retrieval on the related items. These results suggest that inhibition is involved in semantic forgetting.

RIF in Arithmetic

Phenix and Campbell (2004) conducted a study that compared two opposing views of memory for arithmetic facts: the integrated structures model (e.g., Manly & Spoehr, 1999) and RIF (e.g., Anderson, et al., 1994). The integrated structures model assumes that when an item is activated in memory (e.g., 4 × 7 = ?), the items that are related to those operands are activated and strengthened (e.g., 8, 12, 16, 20, 24, 28 and 7, 14, 21 28), and therefore will be easier to retrieve. In contrast, RIF assumes that when an item in memory is activated, the related items are suppressed, which will reduce the speed and accuracy of responses when they are retrieved from memory.

Phenix and Campbell asked undergraduate university students to practice 8 single-digit multiplication problems composed of unique operands for 40 blocks. Unique operand problems are those problems composed of a specific set of operands (i.e., Set A: 2, 5, 7, 8; Set B: 3, 4, 6, 9) and are used to investigate categorical RIF. Categorical RIF occurs for a specific family or category of arithmetic facts. Following this, participants completed a product-verification task where they were asked if presented equations (e.g., 4 × 8 = 24) were true or false. Participants made more errors on related practiced problems, presumably because they shared more common identifiers (e.g., if participants practiced 4 × 7, they were slower to verify the answer to 4 × 8).

The results of this study demonstrated that, consistent with the theoretical assumptions of RIF, retrieval of the multiplication problems caused participants to be slower and make more errors in the verification task, which indicated that when a problem is activated, related items were suppressed and therefore were more difficult to recall. Note that RIF in the standard paradigm is measured in terms of the decreased probability of retrieval of Rp- items, whereas in arithmetic it is expressed in slower response time (RT) or increased errors.

When researchers have investigated RIF of arithmetic facts, they use a modified retrieval-practice paradigm; however, the experimental conditions of the retrieval practice paradigm map onto the paradigm used to investigate RIF of arithmetic facts (e.g., Campbell & Phenix, 2009). The basic phenomena of RIF in arithmetic occurs when practicing retrieval of a specific multiplication counterpart (e.g., 3 × 4 = ?) impairs memory for the addition counterpart (e.g., 3 + 4 = ?). The basic elements of the paradigm included retrieval practice of a subset of simple multiplication problems (e.g., 3 × 4) followed by a test phase in which the addition counterparts (3 + 4) are performed along with control additions whose multiplication counterparts have not been practiced. RIF is expressed in longer RTs or increased errors for addition problems whose multiplication counterparts were practiced (MP additions) compared to addition problems whose multiplication counterparts were unpracticed (MU additions). The multiplication practiced problems (MP) presented in the practice phase correspond to the Rp+ items in the retrieval practice paradigm. The multiplication-unpracticed addition problems (MU) provide a baseline that is equivalent to the Nrp items in the standard RIF paradigm. Finally, the addition problems that correspond to the MP problems are equivalent to the Rp- items, which are items that undergo RIF. It has been consistently demonstrated that RIF occurs when items are retrieved, but not when they are studied (e.g., Campbell and Phenix, 2009; Campbell and Thompson, 2012) and adults primarily retrieve the answers to single-digit multiplication (LeFevre, Bisanz, Daley, Buffone, Greenham, & Sadesky, 1996) and addition problems (LeFevre, Sadesky, & Bisanz, 1996).

Campbell and Thompson (2012) conducted two studies to investigate whether RIF of arithmetic facts is item-specific or only categorical. In the Phenix and Campbell (2004) study, the problems sets assigned to the MP and MU conditions were composed from sets of unique operands (i.e., Set A: 2, 5, 7, 8; Set B: 3, 4, 6, 9) to examine if RIF was categorical (i.e., the 2-times family, the 5-times family, etc. were subject to RIF). Consequently, the RIF observed could be categorical or item specific. Campbell and Thompson composed their problem sets of common operands (i.e., Set A/B: 2, 3, 4, 5, 6, 7, 8, 9) to examine if RIF is item-specific (i.e., practicing 2 × 4 causes RIF of 2 + 4). If arithmetic RIF is only categorical, and not item-specific, then there would be no RIF with this composition of MP and MU problems because all operands 2 through 9 are used as category retrieval cues. In the multiplication practice phase, participants either studied (e.g., 4 × 6 = 24 and state the presented answer) or retrieved (e.g., 4 × 6 = ?) the answers to problems for six blocks. The practice phase was followed by two addition posttest blocks, with an intervening multiplication posttest in Experiment 1, but not in Experiment 2. In both experiments Campbell and Thompson replicated the finding of addition RIF following multiplication retrieval, but not multiplication study. RIF was found when stimuli were composed of common MP and MU operands for the retrieval group, which confirmed that RIF of addition facts is item-specific RIF, although it did not rule out category based RIF. Both experiments also found RIF for small additions problems with a sum less than 10, but not for larger problems. In fact, for larger single-digit additions with sum > 10, Campbell and Thompson's participants showed evidence of priming in that the RT for large MP problems tended to be faster than for large MU problems. Campbell and Thompson argued that RIF only occurred with small problems because North American participants' memory strength is high for small problems, which attracted more inhibition to these problems during multiplication practice, compared to large problems. Campbell and Thompson argued that when RIF does not occur, priming may occur, which can work against RIF. When RIF is observed, it is because the effects of inhibition are stronger than priming and therefore overcomes priming effects.

The different RIF results for small and large problems is related to the problem-size effect, which is the nearly-ubiquitous finding that small simple-arithmetic problems (i.e., sum ≤ 10) (e.g., 2 + 3, 3 × 4) are solved more quickly and accurately than large ones (6 + 7, 7× 8) (see e.g., Campbell, 1995). The primary explanations for why participants solve small problems faster and more accurately than large problems is that participants typically have more practice on small compared to large problems (Hamann & Ashcraft, 1986) and that procedural strategies such as counting are more accurate for small problems leading to more frequent reinforcement of the correct association in memory during initial learning (Siegler, 1988). Campbell and Oliphant (1992) argue that larger problems, for example additions with sums greater than 10, have a less distinctive memory representation because the representation of magnitude is less precise as magnitude increases.

The role of problem size in addition RIF, however, is complicated by results that differed from Campbell and Thompson (2012) (i.e., RIF for small but not large additions). Campbell and Dowd (2012) examined RIF of arithmetic facts for Chinese-English bilingual participants. Specifically, they wanted to investigate if Chinese-English bilinguals had number-fact representations in both English and Chinese. Participants first completed an addition pretest (e.g., 6 + 8) of all 36 problems and responded in either English or Chinese. Next, participants were asked to complete an MP practice phase (e.g., 6 × 8), where they practiced retrieval on a subset of the problems that they solved in the addition pretest (i.e., 6 blocks of 12 problems). Participants responded in English and Chinese, which alternated across the practice blocks and was counterbalanced across participants. Finally, participants completed an addition posttest that was composed of the same problems in the addition pretest (i.e., to measure RIF). RIF was found for large addition problems (i.e., sum of the operands > 10) when participants practiced their multiplication counterparts in the same language but not when the languages of multiplication practice and addition test were different. In contrast to Campbell and Thompson, Campbell and Dowd did not observe RIF for small problems. Thus, the effects of problem-size on RIF reported by Campbell and Dowd were opposite to those reported by Campbell and Thompson because Campbell and Dowd found RIF for large but not small problems.

Campbell and Therriault (2013) proposed that the different effects of problem size on RIF found by Campbell and Dowd (2012) and Campbell and Thompson owed to differences in the stimulus sets used. Specifically, the stimuli used by Campbell and Thompson were entirely fact based (i.e., composed of the operands 2 through 9), while one third of the stimuli used for small problems by Campbell and Dowd were rule based (i.e., problems composed with the operands 0 or 1).Rule-based problems are not susceptible to RIF because these problems are solved using a rule (e.g., N × 1 = N) and are not solved by retrieval of a specific fact from memory. Campbell and Therriault investigated if the rule-governed problems in Campbell and Dowd reduced the observed RIF effects on small problems. First, Campbell and Therriault reanalysed Campbell and Dowd's data on the small problems, excluding the rule-governed problems. When the rule-governed addition problems (i.e., N + 0 = N) were eliminated, RIF was found for the small fact-based problems, but not for the large problems and only when participant's practiced multiplication retrieval in the same language as the addition test.

To examine this further, Campbell and Therriault (2013) conducted a study to examine RIF for rule-based simple multiplication and addition problems. Participants retrieved four blocks of small fact-based (e.g., 2 × 3) and rule- based problems (e.g., 2 × 1; 2 × 0). Following multiplication practice (e.g., 4 × 3 = ?) , participants solved two blocks of the addition counterparts (e.g., 4 + 3 = ?) and control additions (e.g., 5 + 2 = ?) whose counterparts had not been practiced (i.e., both MP and MU problems). Results demonstrated that RIF occurred for the fact-based problems, but there was no evidence of RIF for the rule-based problems. Campbell and Therriault concluded that RIF only occurred for the fact-based problems, but not the rule-based problems because these problems are solved using rules, rather than retrieval of individual facts. Therefore, the inclusion of the rule-based problems by Campbell and Dowd (2012) would have moderated RIF for the small problems. When the rule-based problems were removed and the data reanalyzed, RIF was found for the small problems in the Campbell and Dowd experiment (Campbell & Therriault, 2013).

Hyper-RIF in Arithmetic

Campbell and Phenix (2009) conducted an experiment to examine RIF effects of retrieval practice of single-digit multiplication problems (e.g., 4 × 5) on subsequent retrieval of corresponding addition problems (e.g., 4 + 5). Sixty-six university students first completed two blocks of single-digit addition problems and then either studied (e.g., 4 × 7 = 28) or retrieved (4 × 7 = ?) a subset of multiplication problems for 40 blocks. These were the MP problems. Those problems that the participants did not practice were defined as MU problems. In the subsequent post-practice test phase, participants completed several alternating retrieval blocks of addition and multiplication that included both MP and MU problems.

In the first addition posttest block, RIF was expressed by slower response time (RT) on MP additions compared to MU additions. As shown in Figure 1-1, the retrieval group, but not the study group, presented longer mean RT for the MP relative to MU addition problems in the first addition posttest. This provided evidence that RIF in arithmetic is retrieval dependent, because RIF occurred only for the retrieval group (e.g., Anderson et al., 2000). As retrieval and study practice strengthened the multiplication facts equally, the results supported an inhibition account over an interference account of arithmetic RIF because RIF occurred for the retrieval group, but not the study group (i.e., retrieval dependency).

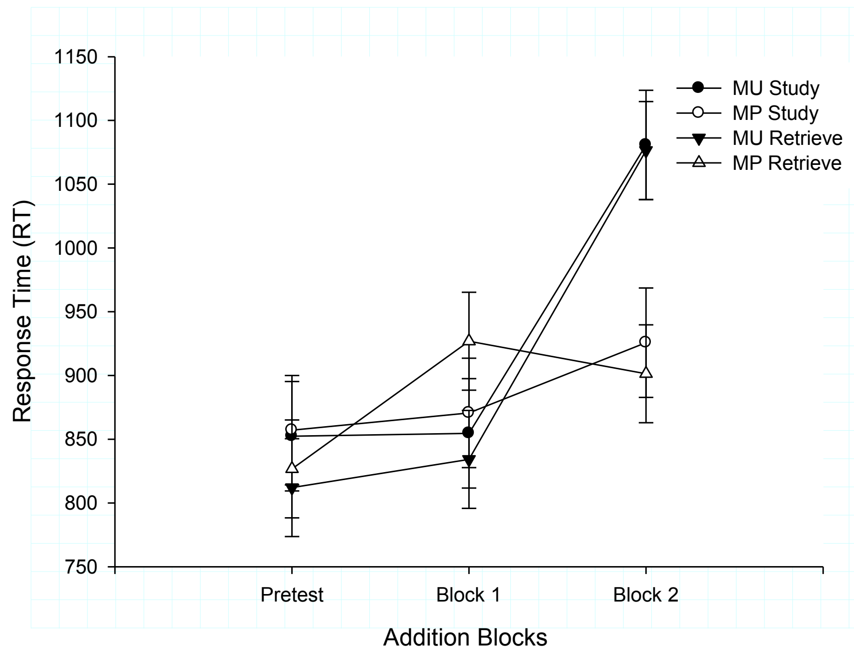
Between the first and second addition posttest block, there was a single multiplication block in which both MP and MU multiplication problems received a single retrieval trial. Interestingly, there was a surprising effect in the second addition test block: participants' RT for MU addition problems increased dramatically from Block 1 to Block 2 for both the retrieval and study groups (see Figure 1-1). RIF in Block 2 for MU addition was about 200% greater than the RIF observed in Block 1 for the multiplication practice retrieval group. This *hyper-RIF* effect occurred presumably because of the intervening multiplication retrieval block, which included both MP and MU problems.

To explain these findings, Campbell and Phenix proposed that 40 repetitions of the MP multiplication problems during the practice phase yielded such high memory strength for these items that, by the end of practice, the addition counterparts were relatively weak retrieval competitors and would not attract inhibition and RIF (see e.g., Storm & Levy, 2012, for a discussion of competition dependence of RIF). Thus, the addition RIF effect observed for the retrieval group in Block 1 was relatively weak. In contrast, a single retrieval of the MU problems in the intervening multiplication block exerted maximum inhibition of the addition counterparts, yielding hyper-RIF for MU items for both groups in addition Block 2. For the MP additions in Block 2, neither the retrieval or study group showed substantial RIF, presumably because both multiplication retrieval and study during the 40 blocks of practice produced very high memory strength for these facts, making their addition counterparts weak competitors and therefore subject to weak RIF.

The experiments reported in this thesis focus on the hyper-retrieval-induced forgetting effect reported by Campbell and Phenix (2009). Hyper-RIF is potentially an important result theoretically. A larger RIF effect with one target retrieval practice trial, compared to a great amount of target retrieval practice, is more consistent with inhibitory accounts of RIF (e.g., Anderson, 2003) than interference accounts (e.g., Butler, William, Zachs, & Maki, 2001), because the latter predict that RIF should increase with target strength (i.e., with target retrieval practice).

Figure 1-1

Hyper-RIF Adapted from Campbell and Phenix (2009, Figure 2). Mean Median Correct Addition Response Times (RT) in milliseconds in Addition Blocks as a Function of Experimental Block, Practice, and Group



Note. Adapted from Campbell and Phenix (2009, Figure 2). Mean correct addition response times (RT's in milliseconds) in addition blocks as a function of experimental block, practice, and group. Error bars are repeated measure 95% confidence intervals (Jarmasz & Hollands, 2009).

Subsequent Failures to Observe Hyper-RIF.

Campbell and Thompson (2012) and Campbell and Phenix (2009), both inserted a multiplication retrieval block that included both MP and MU problems between the first and second addition blocks. This potentially provided an opportunity to observe hyper-RIF. In two experiments, although Campbell and Thompson replicated the finding of addition RIF following retrieval but not multiplication study in the multiplication practice phase in addition Block 1, neither experiment found evidence for hyper-RIF in the second addition posttest block.

The Present Experiments

The questions motivating the current experiments surround the hyper-RIF effect reported by Campbell and Phenix (2009) but not reproduced by Campbell and Thompson (2012). Did hyper-RIF occur for Campbell and Phenix because they used unique operands for the MP and MU sets (Set A: 2, 5, 7, 8; Set B: 3, 4, 6, 9) and many retrieval repetitions of the practice multiplication problems (40 blocks of MP problems), whereas Campbell and Thompson did not replicate the hyper-RIF effect using common operands (Set A/B: 2, 3, 4, 5, 6, 7, 8, 9) and less practice (6 blocks of MP problems)? Experiment 1 manipulated the differences between the Campbell and Phenix and Campbell and Thompson in a 2 (more/less practice) × 2 (unique/common operands) between-participants design. Experiment 2 attempted an identical replication of Campbell and Phenix's original study.

CHAPTER 2

Campbell and Phenix (2009) observed retrieval-induced forgetting (RIF) (slower response time) for simple addition facts (e.g., 3 + 4) immediately following 40 retrieval-practice blocks of their multiplication counterparts (3 × 4 = ?). A subsequent single retrieval of the previously unpracticed multiplication problems, however, produced an RIF effect about twice as large for their addition counterparts. Thus, a single retrieval of a multiplication fact appeared to produce much larger RIF of the addition counterpart than did many multiplication retrieval-practice trials. In subsequent similar studies, however, this *hyper-RIF* effect was not observed (e.g., Campbell & Thompson, 2012). The current studies further investigated hyper-RIF in arithmetic. In Experiment 1, composition of operands (unique, common) and amount of multiplication practice (6 vs. 20 repetitions of each problem) were manipulated. Participants solved multiplication problems (4 × 7 = ?) and then were tested on their memory for the addition counterparts (4 + 7 = ?) and control additions. Three groups demonstrated typical RIF effects and one group did not, but instead demonstrated evidence of priming. No evidence of hyper-RIF was found.

Experiment 1

The present experiment pursued the hyper-RIF effect observed by Campbell and Phenix (2009), but not by Campbell and Thompson (2012). Using E-Prime 2.0, participants were asked to retrieve the answers to the presented problems (both MP and MU) and speed and accuracy of each response was recorded. We wanted to determine if RIF affects the memory performance of individual addition facts or entire categories of facts in memory. Isolating the conditions that produce hyper-RIF will accordingly provide the foundation for the development of a theoretical model to help explain this important phenomenon.

Method

Participants

Seventy-two undergraduate psychology students (16 males) at the University of Saskatchewan, Saskatchewan, Canada completed this experiment for extra course credit. Participants were randomly assigned evenly across the four groups [i.e., less practice, unique operands (4 males); less practice, common operands (3 males); more practice, unique operands (5 males); more practice, common operands (4 males), respectively (n = 18; n = 18; n =18; n = 18)]. Age of participants ranged from 17 to 25 (M = 18.81).

Stimuli and Design

Experiment 1 used a combination of both Campbell and Phenix (2009) and Campbell and Thompson's (2012) experimental designs. A total of 16 addition problems (and their multiplication counterparts) were used. Two problem sets (common and unique) were broken down into MP and MU problems and counterbalanced across participants. MU problems were used as control problems. Unique problem sets were composed of the operands 2, 5, 7, 8 and 3, 4, 6, 9. Both common operand sets were composed of the operands 2 through 9. The unique-operand sets were taken directly from Campbell and Phenix (2009). Set 1 included the problems

2 + 5 = 7, 7 + 2 = 9, 8 + 2 = 10, 5 + 5 = 10, 5 + 7 =12, 8 + 5 = 13, 7 + 7 = 14, 7 + 8 = 15 and Set 2 included the problems 3 + 4 = 7, 6 + 3 = 9, 4 + 4 = 8, 4 + 6 = 10, 9 + 3 = 12, 9 + 4 = 13, 6 + 6 = 12, 6 + 9 = 15. Using these unique sets, two common operand problem sets were created, which combined both unique problem sets so that operands 2 through 9 were used: 2 + 5 = 7, 8 + 2 = 10, 5 +7 = 12, 7 + 7 = 14, 3 + 4 = 7, 4 + 4 = 8, 9 + 3 = 12, 6 + 9 = 15, 7 + 2 = 9, 5 + 5 = 10, 8 + 5 = 13, 7 + 8 = 15, 6 + 3 = 9, 4 + 6 = 10, 9 + 4 = 13, 6 + 6 =12. During multiplication blocks, participants solved the corresponding multiplication counterparts of these problems. The unique sets correspond to a unique category or family of operands in memory. The common sets used all possible operands between 2 through 9 and therefore do not correspond to a specific family of arithmetic facts.

The experiment was comprised of an addition pretest, an MP practice phase, one multiplication and two addition posttests. In the addition pretest, all participants solved one block of 16 addition problems. In the MP practice phase, participants either practiced 6 or 20 blocks of multiplication of 8 MP problems that were counterbalanced across groups. Campbell and Phenix (2009) used 40 practice blocks and we used 20 in our experiment. We expected that 20 practice blocks would be enough retrieval practice to elicit high target strength. Participants solved one block of the 16 addition problems again that contained both MP and MU problems, followed by one block of their corresponding multiplication counterparts that also contained both MP and MU problems, and a final block of the addition posttest. The addition pretest and posttests were comprised of the same 16 problems. All participants were asked to retrieve the answers to the problems (e.g., 5 + 7 = ?). Retrieval but not study of arithmetic facts was used in Experiment 1.

With the combination of problem set (common vs. unique operands) and number of practice blocks (6 vs. 20), participants were assigned to one of four possible groups. We used abbreviations for our between-participants factors. OpsC is the abbreviation for the common (i.e., two sets consisting of the operands 2 through 9) and OpsU is the abbreviation for the unique set (i.e., two sets consisting of operands 2, 5, 7, 8 or 3, 4, 6, 9). The abbreviation MP6/20 is used for the amount multiplication practice blocks, because participants either completed 6 or 20 blocks of multiplication practice. MpMu is our abbreviation for the factor of multiplication practice where MP is practiced and MU is unpracticed.

In all three experimental phases, problems were presented in white font against a black background that spanned the length of 5 characters in point 18 Courier New font (e.g., 2 × 3).

Procedure

The experiment took place in a quiet room and took 30 minutes to complete. Prior to beginning the experiment, participants signed a consent form which indicated their willingness to participate. Problems were presented to the participant using a high resolution cathode ray tube (CRT) monitor. Instructions were given before each block of problems and indicated to the participant to respond as quickly and accurately as possible. Participants responded verbally using a microphone that controlled a software clock accurate to ± 1 millisecond. First, participants completed an addition pretest that was comprised of all 16 problems, followed by a practice phase where they practiced 8 MP multiplication problems (described above) for 6 or 20 blocks. Participants were told that they would be tested on the MP practice questions later on in the experiment. Next, participants completed an addition posttest, followed by a test of their multiplication counterparts, and finally completed a second addition posttest. All addition blocks were comprised of both the MP and MU problems and the multiplication posttest was composed of the corresponding multiplication counterparts.

Before the beginning of each trial, a fixation dot appeared in the center of the screen for one second followed by the problem that was presented as an addition (e.g., 3 + 4) or a multiplication problem (3 × 4). RT and accuracy of each trial was recorded.

On a separate monitor, the experimenter kept track of participants' responses and marked incorrect and spoiled trials (i.e., if the microphone timing was inaccurate). The experimenter began each block of arithmetic problems by pressing the spacebar. Throughout the experiment, participants were encouraged to respond as quickly and accurately as possible.

Across the four different groups [OpsU/6; OpsU/20 (Campbell & Phenix, 2009 replication group); OpsC/6 (Campbell & Thompson, 2012 replication group); OpsC/20], we expected RIF to occur during addition Block 1 for MP problems, relative to MU, because all four groups retrieved MP problems. We expected that the Campbell and Phenix replication group would yield similar results to Campbell and Phenix's original study: RIF should occur for MP, relative to MU problems in addition Block 1 and evidence of hyper-RIF in addition posttest Block 2 (i.e., a single retrieval of MU problems should inhibit retrieval of their addition counterparts and produce an exaggerated RIF effect relative to Block 1). We also expected to replicate the original findings of Campbell and Thompson in our replication group. Specifically, RIF would occur for small problems and priming would occur for large problems, but there would be no evidence of hyper-RIF.

It should be noted that Campbell and Phenix's (2009) stimuli were not originally intended to be broken down by size (small and large), but we added size as a within participant factor in order to compare our Campbell and Phenix replication group with the Campbell and Thompson (2012) replication group, whose original stimuli were broken down by size. It was decided that it was best to analyze median RT, rather than mean RT because of the small number of observations per cell. Medians are less subject to the influence of outliers than means.

Results

As outlined in the procedure section, participants completed blocks of multiplication problems followed by two addition posttests, with an intervening multiplication posttest. Participants were randomly assigned to conditions including operands and amount of multiplication practice. The sum of operands was ≤ 10 for small problems (Campbell & Thompson, 2012; LeFevre et al., 1996) (i.e., size). A total of 507 RTs (3.0%) were marked by the experimenter as spoiled and therefore were not included in the analyses. The purpose of the analyses was to determine if there were differences based on the respective between or within factors described above, as well as to determine if there were interactions among them. First we presented results from the multiplication practice and test phase followed by analyses of the addition pretest and posttest. Participant's raw scores entailed calculating the median response for RT. The different analysis of variances (ANOVAs) entailed conducting an analysis of the means of these median scores. Significance tests with $p \leq .10$ are reported (see Appendix A).

Multiplication Practice

RT.

RT analysis in the multiplication practice phase was conducted to determine the degree of strengthening differences in the multiplication practice across set and operand type. Early and late practice blocks were defined as the first three practice blocks versus the last three blocks, respectively, for both MP 6 and 20 practice blocks. A repeated measures ANOVA was conducted on mean of the median (hereafter, referred to as the mean median) correct RT using a 2 (Size; Small/Large)[1] × 2 (Block; Early/Late) × 2 (OpsC/U) × 2 (MP6/20) design. Between-participant factors were OpsC/U and MP6/20 and within-participant factors were block and size. The results of this analysis appear in Table 2-1.

Participants were faster on small problems (800 ms) than large problems (916 ms), which is the standard effect of problem size [F (1, 68) = 42.256, MSE = 22815.443, $p < .001$, ηp^2 = .383]. Participants were faster late in practice (786 ms) compared to early practice (930 ms) [F (1, 68) = 41.988, MSE = 35465.028, $p < .001$, ηp^2 = .382]. As would be expected, participants who solved 20 MP blocks sped up more (-197 ms) than those who solved 6 MP blocks (-91 ms). Thus, as would be expected, more repetitions of each problem lead to faster RTs [F (1, 68) = 5.746, MSE = 35465.028, p = .019, ηp^2 = .078].

There was weak evidence for a Size × Block × MP6/20 interaction [F (1, 68) = 3.761, MSE = 16062.734, p = .057, ηp^2 = .052]. Specifically, early in practice, participants who received 20 blocks of practice were faster on small than large problems (+180 ms), compared to participants who received 6 blocks of practice (+76 ms), but late in practice the problem size effect was more similar for participants who received 6 (+110 ms) and 20 (+98 ms) blocks of practice.

Finally, there was a Size × Block × OpsC/U interaction [F (1, 68) = 5.367, MSE = 16062.734, p = .024, ηp^2 = .073]. Early in practice, participants who solved the unique-operand problem set had a smaller problems size effects (+74 ms) than participants who solved the common-operand set (+181 ms). In late practice, this pattern reversed, with a trend for a larger problem size effect for the unique-operand set (+120 ms) than the common-operand set (+88 ms). In general, one would expect the problem size effect to decrease with practice because the slower, large problems will benefit more from each practice trial. Thus, the common operand groups' much-reduced problem size effect late in practice corresponded to the expected effects of practice. The increased problem size effect late in practice for the unique operands groups reflected relatively little speed up for the large problems compared to the small problems, especially for the OpsC/6 group. There was no related effect in the analysis of errors in the multiplication practice phase (see below). Thus, different speed accuracy trade off criteria across groups cannot explain the unusual multiplication RT results. We do not have an explanation for this unexpected result.

Table 2-1

Experiment 1 Mean Median Response Time (RT) in milliseconds across Groups by Size in the Multiplication Practice Phase and Posttest

	Group			
	OpsU/6	OpsU/20	OpsC/6	OpsC/20
Early Multiplication Practice				
Small	872 (62.8)	949 (78.2)	821 (40.1)	820 (29.3)
Large	896 (66.0)	1074 (107.3)	949 (58.4)	1054 (68.1)
Late Multiplication Practice				
Small	732 (36.5)	797 (48.8)	747 (46.0)	661 (19.2)
Large	870 (79.4)	897 (67.6)	828 (53.5)	755 (37.2)
Multiplication Posttest				
Small MP	866 (72.5)	890 (62.5)	843 (38.6)	779 (25.5)
Large MP	901 (76.1)	1002 (70.7)	919 (46.4)	951 (65.1)
Small MU	955 (61.3)	1283 (122.2)	1021 (46.4)	1001 (57.3)
Large MU	1144 (114.0)	1377 (100.0)	1233 (98.5)	1133 (90.7)

Note. Standard error in brackets. OpsU/6 = unique operands, less practice; OpsU/20 = unique operands, more practice; OpsC/6 = common operands, less practice, OpsC/20 = common operands, more practice; MP = multiplication practiced; MU = multiplication unpracticed.

Percent Error.

Percentage of errors in the multiplication practice phase was conducted to determine if there were differences in the amount of errors made across set and operand type. Percent error was analyzed using a 2 (Size; Small/ Large) × 2 (Block; Early/Late) × 2 (OpsC/U) × 2 (MP6/20) design. Between-participant factors were OpsC/U and MP6/20, and block and size were within-participant factors. Participants made fewer errors on small (2.8%) than large problems (12.1%) [F (1, 68) = 38.242, MSE = 163.441, p < .001, ηp^2 = .360], and made more errors early (11.1%) compared to late in practice (3.8%) [F (1, 68) = 52.258, MSE = 72.097, p < .001, ηp^2 = .435]. There was a Size × Block interaction whereby from early to late practice, participants' errors decreased more on large problems (12.8%) than small problems (1.6%) [F (1, 68) = 36.898, MSE = 61.487, p < .001, ηp^2 = .352] (see Table 2-2).

Table 2-2

Mean Error Rates (%E) in the Multiplication Practice Phase

	Early	Late
Small MP	3.6 (0.7)	2.0 (0.5)
Large MP	18.5 (2.2)	5.7 (1.3)

Multiplication Posttest

RT.

A repeated measures ANOVA was conducted on mean median correct RT in the multiplication posttest using a 2 (Size; Small/ Large) × 2 (multiplication practice vs. unpracticed, henceforth MpMu) × 2 (OpsC/U) × 2 (MP6/20) design. This allowed us to confirm that practice of MP problems in the practice phase resulted in better performance compared to the MU problems. One participant had no correct answers on MU large problems and was excluded from the analysis. The means appear in Table 2-1. For the remaining participants, there was an effect of MpMu whereby participants were faster on MP problems (894 ms) than MU problems (1144 ms) [F (1, 67) = 101.595, MSE = 43513.713, p < .001 ηp^2= .603]. It is clear that the effect of multiplication practice in the practice phase was effective because participants were faster on the MP problems in the multiplication posttest. There was an effect of size whereby participants showed evidence of the problem size effect [F (1, 67) = 17.424, MSE = 665256.765, p < .001,

$\eta p^2 = .206$] because they were faster on small (955 ms) than large problems (1083 ms). The MpMu × MP6/20 interaction approached significance [F (1, 67) = 3.107, MSE = 43513.713, p =.082, $\eta p^2 = .044$] whereby participants who received 20 blocks of practice were faster on MP than MU problems (-293 ms), compared to those who received 6 blocks of practice, who were also faster on MP problems, but by a smaller margin (-206 ms). This would be expected because the those participants had 20 blocks of multiplication practice, had more practice and therefore were faster than those who solved six blocks of multiplication practice. A MpMu × OpsC/U × MP6/20 interaction occurred [F (1, 67) = 7.041, MSE = 4353.713, p = .010, $\eta p^2 = .095$]. Specifically, with six practice blocks, there was a greater RT advantage for MP relative to MU problems with common operands (-246 ms) than unique operands (-166 ms). In contrast, with 20 practice blocks, the MP advantage tended to be smaller with common operands (-202 ms) than unique operands (-385 ms). This was unexpected and we currently have no explanation as to why this occurred. Most important though, participants were faster to solve MP problems than MU problems, which suggests that the multiplication practice phase was effective. This is important for typical RIF effects to occur in the addition posttest blocks.

Participants slowed down slightly in the multiplication posttest, compared to RT late in the MP practice phase (see Table 2-1). This most likely reflected interference created by the first addition posttest block that participants completed just prior to the multiplication posttest. This change from solving multiplication, to addition, back to multiplication may have caused participants to slow down on the multiplication posttest. The same pattern of results occurred in Campbell and Phenix’s (2009) original study.

Percent Error.

Percentage of errors in the multiplication posttest phase was also analyzed using a 2 (Size; Small/ Large) × 2 (MpMu) × 2 (OpsC/U) × 2 (MP6/20) design. The means appear in Table 2-3. As expected, participants made fewer errors on small (1.1%) than large problems (3.2%) [$F\,(1, 68) = 28.928$, $MSE = 10.083$, $p < .001$, $\eta p^2 = .298$], and made less errors on MP (0.8%) than MU (3.4%) [$F\,(1, 68) = 49.275$, $MSE = 9.909$, $p < .001$, $\eta p^2 = .420$]. There was a Size × MpMu interaction whereby, participants' made more errors on large than small MU problems (+3.0) and made more errors on large than small MP problems, but the difference was not as large (+1.1%) [$F\,(1, 68) = 6.438$, $MSE = 10.196$, $p = .013$, $\eta p^2 = .086$]. The results of the percentage of errors multiplication posttest were consistent with our expectations. Participants made fewer errors on small than large problems and fewer errors on MP compared to MU problems. The fact that participants made fewer errors on MU problems with 20 blocks of practice compared to 6 blocks of practice was expected because participants who completed 20 blocks of practice had more practice than those with 6 blocks of practice.

Table 2-3

Experiment 1 Mean Error Rates (%E) across Groups by Size in the Multiplication Posttest

	Group			
	OpsU/6	OpsU/20	OpsC/6	OpsC/20
Small MP	0.4 (0.3)	0.0 (0.3)	0.3 (0.3)	0.3 (0.3)
Large MP	2.1 (0.8)	0.7 (0.8)	2.1 (7.5)	0.7 (0.8)
Small MU	2.4 (0.8)	2.1 (0.8)	1.7 (8.2)	1.4 (0.8)
Large MU	4.5 (1.1)	4.9 (1.1)	5.2 (1.1)	5.2 (1.1)

Note. Standard error in brackets. OpsU/6 = unique operands, less practice; OpsU/20 = unique operands, more practice; OpsC/6 = common operands, less practice, OpsC/20 = common operands, more practice; MP = multiplication practiced; MU = multiplication unpracticed.

Addition Pretest

RT.

The addition pretest was analyzed to ensure there were no group differences prior to the multiplication practice phase. A 2 (Size; Small/ Large) × 2 (MpMu) × 2 (OpsC/U) × 2 (MP6/20) repeated measures ANOVA on mean median correct RT in the addition pretest was conducted. The means appear in Table 2-4. As expected, there was evidence of a problem size effect as participants were faster on small (795 ms) than large problems (1023 ms) [$F(1, 68) = 64.328$, $MSE = 58032.948$, $p < .001$, $\eta p^2 = .486$]. As expected, there were no RT differences between participants on the addition pretest (all group related effects had $p > .10$).

Table 2-4

Experiment 1 Group by Size and MpMu Mean Median Response Time (RT) in milliseconds in the Addition Pretest

	Group			
	OpsU/6	OpsU/20	OpsC/6	OpsC/20
Pretest	RT	RT	RT	RT
Small				
MP	774 (56.2)	843 (46.2)	826 (45.7)	725 (18.8)
MU	798 (63.3)	808 (35.4)	833 (61.4)	756 (31.5)
Large				
MP	981 (85.5)	1051 (82.2)	1025 (67.5)	981 (39.9)
MU	1033 (85.0)	1102 (83.4)	1092 (108.1)	919 (44.0)

Note. Standard error in brackets. OpsU/6 = unique operands, less practice; OpsU/20 = unique operands, more practice; OpsC/6 = common operands, less practice, OpsC/20 = common operands, more practice; MP = multiplication practiced; MU = multiplication unpracticed.

Percent Error.

The same ANOVA was conducted on percentage of error in the addition pretest to make sure there were no differences group differences prior to multiplication practice. The means appear in Table 2-5. An effect of size was observed [$F(1, 68) = 20.051$, $MSE = 14.313$, $p < .001$, $\eta p^2 = .228$]. Participant's made fewer errors on small problems (0.9%) than large problems (2.9%), which again reflect the problem size effect. As expected, there were no error differences between the groups in the addition pretest.

Table 2-5

Experiment 1 Mean Error Rates (%E) across Groups by Size in the Addition Pretest

	Group			
	OpsU/6	OpsU/20	OpsC/6	OpsC/20
Small MP	1.7 (0.6)	0.3 (0.6)	1.4 (0.6)	0.3 (0.6)
Large MP	4.5 (1.1)	2.4 (1.1)	3.5 (1.1)	2.1 (1.1)
Small MU	0.7 (0.6)	2.1 (0.6)	0.3 (0.6)	0.3 (0.6)
Large MU	3.5 (0.9)	2.4 (0.9)	3.1 (0.9)	1.7 (0.9)

Note. Standard error in brackets. OpsU/6 = unique operands, less practice; OpsU/20 = unique operands, more practice; OpsC/6 = common operands, less practice, OpsC/20 = common operands, more practice; MP = multiplication practiced; MU = multiplication unpracticed.

Addition Posttests

RT.

Analyses of the addition posttests provided the opportunity to investigate RIF effects. Block 1 of the addition posttest provided the purest measure of effects of multiplication practice on MP and MU addition (i.e., RIF), whereas Block 2 provided a potential measure of hyper-RIF induced by the inter-leaved block of multiplication that included both MP and MU problems. It was expected that hyper-RIF would occur for the Campbell and Phenix (2009) replication group, OpsU20, but not for the other groups. We expected to find typical RIF effects for all groups in addition Block 1. Separate 2 (Size; Small/ Large) × 2 (MpMu) × 2 (OpsC/U) × 2 (MP6/20) repeated measure ANOVAs of mean median correct RT were conducted for each addition posttest block. The means appear in Table 2-6. In Block 1, small problems (819 ms) were solved faster than large problems (977 ms) [$F(1, 68) = 46.895$, $MSE = 38496.710$, $p < .001$, $\eta p^2 = .408$]. Participants were slower to solve MP problems (922 ms) than MU problems (873 ms) [$F(1, 68) = 6.082$, $MSE = 28382.772$, $p = .016$, $\eta p^2 = .082$], which confirms RIF for MP problems in Block

1. Although we had no specific expectations for OpsC/U in Block 1, an MpMu × OpsC/U interaction occurred because RIF occurred for MP problems in the common-operand set (+89 ms), but this did not occur with unique operands (+8 ms) [F (1, 68) = 4.159, MSE = 28382.772, p = .045, $\eta p^2 = .058$]. This effect was qualified, however, by the three-way interaction of MpMu × OpsC/U × MP6/20. Specifically, participants who solved six blocks of MP problems showed RIF for both unique-operand (+82 ms) and common-operand problem sets (+66 ms),and RIF also occurred for participants who solved 20 blocks of MP common-operand problems (+113 ms), but participants who solved 20 blocks of MP unique-operand problems demonstrated significant priming (-65 ms) [F (1, 68) = 5.909, MSE = 28382.772, p = .018, $\eta p^2 = .080$].

Table 2-6

Experiment 1 Mean Median Response Time (RT) in milliseconds across Groups by Size and MpMu in the Addition Posttests

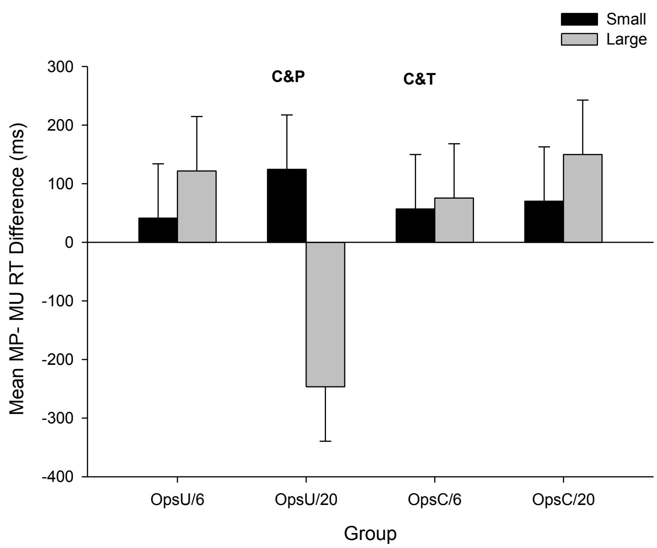
		Group			
	`	OpsU/6	OpsU/20	OpsC/6	OpsC/20
Block 1					
Small					
	MP	811 (62.8)	922 (61.0)	862 (55.5)	822 (30.2)
	MU	770 (60.8)	805 (39.3)	805 (43.8)	752 (31.6)
Large					
	MP	1036 (113.5)	870 (48.5)	1024 (64.5)	1031 (79.4)
	MU	914 (90.0)	1116 (122.4)	949 (67.5)	876 (37.0)
Block 2					
Small					
	MP	818 (65.2)	855 (56.4)	806 (44.1)	795 (37.1)
	MU	857 (87.3)	818 (59.3)	797 (55.2)	757 (45.0)
Large					
	MP	957 (95.8)	981 (62.6)	958 (59.8)	1000 (64.7)
	MU	916 (74.6)	996 (81.9)	885 (48.3)	925 (79.2)

Note. Standard error in brackets. OpsU/6 = unique operands, less practice; OpsU/20 = unique operands, more practice; OpsC/6 = common operands, less practice, OpsC/20 = common operands, more practice; MP = multiplication practiced; MU = multiplication unpracticed.

The Size × MpMu × OpsC/U, a Size × MpMu × MP6/20 and a MpMu × OpsC/U × MP6/20 interactions were significant (all < .05), but these three-way interactions were qualified by a four-way MpMu × Size × OpsC/U × MP6/20 interaction [F (1, 68) = 7.592, MSE = 38813.836, p = .008, ηp^2 = .1]. The four-way effect is shown in Figure 2-1, which plots mean MP RT – mean MU RT differences (positive differences correspond to RIF).

Figure 2-1.

Experiment 1 Addition Block 1 Response Time (RT) RIF across Groups in milliseconds



As Figure 2-1 shows, the Campbell and Phenix (2009) replication group was the source of the four-way interaction. An analysis that included only the other three groups (i.e., OpsU/MP6, OpsC/MP6, and OpsC/MP20) confirmed an overall RIF effect, whereby participants were 87 ms slower on MP (931 ms) than MU (844 ms) [F (1, 51) = 12.422, MSE = 32788.123, p < .001, ηp^2 = .196]. There was no Group × MpMu interaction (p = .734), MpMu × Size interaction (p = .195) or Group × Size × MpMu interaction (p = .808). Thus, there was no evidence that RIF differed among these three groups or that RIF in Block 1 for these groups differed across problem size. This implies that the four-way effect arose in in connection with

OpsU/20. A separate analysis of the OpsU/20 group (i.e., Campbell and Phenix replication group) showed a main effect of size [F (1, 17) = 5.138, MSE = 55205.426, p = .037, $\eta p^2 = .232$], whereby participants were faster on small (867 ms) than large problems (993 ms). The main effect of MpMu approached significance [F (1, 17) = 4.278, MSE = 15698.537, p = .054, $\eta p^2 =$ .201], whereby participants were slightly faster overall on MU problems (900 ms) compared to MP problems (961 ms). As shown in Figure 2-1, however, OpsU/20 presented a robust MpMu × Size interaction [F (1, 17) = 9.468, MSE = 65437.249, p = .007, $\eta p^2 = .358$]. T-tests confirmed that for OpsU/20 there was RIF for small problems (124 ms) [t (16) = 3.063, SE = 40.626, p = .007] whereas the large problems presented priming (-247 ms) [t (16) = 3.216, SE = 76.791, p = .005]. Thus, the four-way effect on RT in Block 1 occurred because only OpsU/20 showed a priming effect of multiplication practice on addition RT and only for large problems.

For addition Block 2, analysis was conducted to examine the hyper-RIF effect. The results of this analysis appear in Table 2-6. A 2 (Size; Small/Large) × 2 (MpMu) × 2 (OpsC/U) × 2 (MP6/20) repeated measures ANOVA of median correct RT demonstrated a main effect of size [F (1, 68) = 54.848, MSE = 25449.211, p < .001, $\eta p^2 = .446$], whereby mean RT was 813 ms for small problems and 952 ms for large problems. The main effect of MpMu approached significance [F (1, 68) = 3.208, MSE = 16749.590, p = .078, $\eta p^2 = .045$] with participants tending to be slower on MP (896 ms) compared to MU problems (869 ms), which suggests that there was a trend for RIF in Block 2, but the effect was not significant. No other effects were found (p's > .10). There was no evidence that hyper-RIF occurred for any of the groups, as indicated by no significant interactions of MpMu with the factors OpsC/U or MP6/20 (all p's > .38).

Percent Error.

To examine group differences in percent error, a separate 2 (Size; Small/ Large) × 2 (MpMu) × 2 (OpsC/U) × 2 (MP6/20) mixed measures ANOVA on percent error was conducted on each addition posttest block. The means appear in Table 2-7. In addition Block 1, there was an effect of size. Specifically, participants made more errors on large (3.2%) than small problems (2.0%) [F (1, 68) = 6.061, MSE = 15.914, p = .016, $\eta p^2 = .016$]. There was also a MpMu × Size × MP6/20 interaction [F (1, 68) = 5.726, MSE = 78.175, p = .019, $\eta p^2 = .078$)]. Participants who received six blocks of practice made more errors on small than large MP problems (1.9%), but made more errors on large than small MU problems (1.6%). In contrast, participants with 20 blocks of practice made less errors on small MP (0.7%) and small MU problems (1.5%). As the differences underlying this interaction are very small, and the pattern makes no apparent sense, there will be no attempt to provide an explanation for this effect. The same analysis on percent error was conducted for addition Block 2, but no effects were found (p's > .10).

Table 2-7

Experiment 1 Error Rate (%E) across Groups by Size and MpMu in the Addition Posttests

		Group			
		OpsU/6	OpsU/20	OpsC/6	OpsC/20
Block 1					
Small					
	MP	3.704 (1.4)	2.315 (1.1)	2.78 (1.0)	1.85 (1.1)
	MU	1.389 (0.8)	1.38 (0.8)	1.389 (0.8)	1.389 (0.8)
Large					
	MP	3.241 (1.4)	3.704 (1.2)	2.315 (0.9)	3.241 (1.2)
	MU	4.167 (1.2)	1.852 (1.1)	4.630 (1.4)	2.315 (0.9)
Block 2					
Small					
	MP	1.852 (0.8)	2.319 (0.9)	.9259 (0.6)	1.852 (0.8)
	MU	2.778 (1.0)	1.852 (0.8)	1.852 (0.8)	2.315 (0.9)
Large					
	MP	2.315 (0.9)	2.315 (0.9)	2.315 (0.9)	1.389 (0.8)
	MU	3.704 (1.4)	3.241 (1.5)	4.167 (1.5)	1.389 (0.8)

Note. Standard error in brackets. OpsU/6 = unique operands, less practice; OpsU/20 = unique operands, more practice; OpsC/6 = common operands, less practice, OpsC/20 = common operands, more practice; MP = multiplication practiced; MU = multiplication unpracticed.

Discussion

Contrary to Campbell and Phenix's (2009) results, Experiment 1 found no evidence of hyper-RIF in any of the experimental conditions. The common/less practice (Campbell & Thompson, 2009 replication group), common/more practice, and unique/less practice groups all presented standard RIF effects in addition RT in Block 1 that did not differ as a function of problem size. In contrast, the OpsU/20 group (the Campbell and Phenix replication group) showed no evidence of hyper-RIF, but instead presented an overall trend for RT priming (see Table 2-6). Table 2-6 and Figure 2-1 show that the priming appeared entirely in connection with the large problems, while small problems showed the standard RIF effect in both Blocks 1 and 2. Thus, the four-way interaction in RT occurred because only the OpsU/20 group showed priming (i.e., a reversal of RIF) and only for the large problems whereas all other Group × Size conditions showed an effect in the direction of RIF (see Figure 2-1).

Why would the OpsU/20 group alone present priming for the large additions? Table 2-6 shows that the OpsU/20 group had particular difficulties with the large addition problems they encountered in the MU condition, although the problems sets were counterbalanced between the MP and MU conditions. It may be that there were pre-experimental differences for participant's memory strength for the large MU problems. The average RT for the OpsU/20 group for the large MU problems was 1116 ms compared to 913 ms on average for the other three groups (i.e., 18.2% slower). On the other hand, the OpsU/20 group was 15.7% faster on the large MP additions (870 ms) compared to the other groups (1030 ms on average). This suggests that this is genuine priming effect from multiplication practice and not just an artifact of long RTs for the large MU additions. Two one-way ANOVAs were conducted comparing the Campbell and Phenix (2009) replication group to the other three groups for RT on large MP and MU problems.

For the large MU problems, the Campbell and Phenix replication group was significantly slower on the large MU problems (1116 ms), compared to the other three groups on average (913 ms) [F (70) = 4.369, p = .040]. For the large MP problems this effect is approaching significance, where the Campbell and Phenix replication group was slightly faster (870 ms) than the other three groups, on average (1031 ms) [F (70) = 3.118, p = .082]

If we assume that their mean RT for the MU problems provides an estimate of RT for large additions in the absence of priming, then their memory strength for large additions (including the MP additions) was quite low compared to the other groups, although this difference was not apparent in the addition pretest. In this case, practice of the large MP multiplication problems would not encounter competition from the large MP additions, and therefore the latter would not attract inhibition and RIF. In the absence of RIF we would expect multiplication practice to prime the addition counterparts, as others have observed (Campbell & Dowd, 2012; Campbell & Thompson, 2012). There were no clear, consistent effects of the between-participant factors (i.e., unique vs. common operands for the MP and MU sets and 6 vs. 20 blocks of multiplication practice) on RIF in this paradigm. Moreover, the two replications groups (i.e., OpsU/20 was a partial replication of Campbell & Phenix, 2009 and OpsC/6 was partial replication of Campbell and Thompson, 2012) did not produce the same RIF results as the original studies by Campbell and Phenix and Campbell and Thompson. Campbell and Thompson observed RIF in RT for small additions but not large additions; whereas their OpsC/6 replication group produced RIF for both small and large additions, as did OpsC/20 and also OpsU/6. Campbell and Phenix did not manipulate or analyze problem size, but their replication group produced the effects of problem size on addition RIF observed by Campbell and Thompson (i.e., RIF for small problems with a trend for priming of large problems).

RIF in addition presumably depends on the memory strengths of corresponding addition and multiplication problems in order for there to be sufficient retrieval competition during multiplication and sufficient memory strength to support addition fact retrieval. To assess if the Campbell and Phenix (2009) and Campbell and Thompson (2012) replication groups were similar in basic multiplication and addition skills to the original participants Table 2-8 was prepared to afford comparisons and the data were collapsed over problem size and only the retrieval-practice groups from the original studies are presented. The original groups made more errors, especially in the first posttest block, but were similar in average RT, except for MU problems in Block 2 of the Campbell and Phenix original, which reflects the sought-after hyper-RIF that was not replicated in Experiment 1. For the other cells, the differences between the performance of the original participants in Campbell and Phenix and Campbell and Thompson studies and the present replication groups seem small enough to discount as likely sources of the different results. Therefore, Experiment 1 showed evidence for the typical RIF in three groups (OpsU/6. OpsC/6, and OpsC/20), but there was no evidence of hyper-RIF in any of the groups and failure to find the hyper-RIF is unlikely to have occurred because of inherent differences in our sample, compared to those who participated in the Campbell and Phenix and Campbell and Thompson studies.

Table 2-8

Experiment 1 Comparison of Percent Error (%E) and Response time (RT) in milliseconds of Campbell and Phenix (2009), Campbell and Thompson (2012), and Replication Retrieval Groups of MpMu in the Addition Posttests

	Group							
	C&P Replication		C&P Original		C&T Replication		C&T Original	
Block 1	RT	%E	RT	%E	RT	%E	RT	%E
MP	896	2.3	927	5.0	943	1.9	935	4.7
MU	961	1.2	834	8.5	877	2.3	915	2.7
Block 2	RT	%E	RT	%E	RT	%E	RT	%E
MP	918	1.7	901	1.9	882	1.2	895	4.1
MU	907	1.9	1076	2.1	841	2.3	865	3.1

Note. MP = multiplication practiced; MU = multiplication unpracticed. C&P = Campbell and Phenix (2009). C&T = Campbell and Thompson (2012)

CHAPTER 3

In Experiment 2 we attempted an exact replication of Campbell and Phenix, but hyper-RIF was not observed. The results confirmed the basic RIF effect of multiplication retrieval practice on addition counterparts, but cast doubt on the on the reality of the hyper-RIF effect observed by Campbell and Phenix. It is concluded that the hyper-RIF effect reported by Campbell and Phenix is an elusive or non-existent phenomenon; consequently, it cannot at this time be considered an important result in the RIF literature.

This chapter has previously been published in an American Psychological Association journal with the Canadian Psychological Association:

Maslany, A. J., & Campbell, J. I. D., (2013), Failures to replicate hyper-retrieval-induced forgetting in arithmetic facts. *Canadian Journal of Experimental Psychology, 67*, 72-77. doi: 10.1037/a0031138

Experiment 2

The results of Experiment 1 indicate that the basic addition RIF effect observed in Block 1 is robust, although there was inconsistency in the appearance of the effect for both small and large simple addition problems. Experiment 1 did not provide any evidence of the hyper-RIF effect reported by Campbell and Phenix (2009) but it was not an identical replication. For example, the Campbell and Phenix group in Experiment 1 used 20 practice blocks, whereas Campbell and Phenix used 40 practice blocks. As the hyper-RIF effect was the primary motivation for Experiment 1, and it failed to produce hyper-RIF, the next step is to attempt an as-close-as-possible direct replication of Campbell and Phenix's original study. Therefore we used the same stimuli and procedure reported by Campbell and Phenix. We expected that the hyper-RIF effect would occur in a close replication of Campbell and Phenix's original experiment.

Method

Participants

Sixty undergraduate psychology students (15 males) at the University of Saskatchewan completed the experiment for bonus marks in their introductory psychology course. Alternating participants were assigned to the retrieval group or study group (both $n = 30$). Males and females were randomly assigned to either the retrieval (8 males) or study group (7 males). Age of participants ranged from 18 to 33 ($M = 20.25$).

Stimuli and Design

Stimuli were presented on two high-resolution monitors using Eprime 2.0 (Psychology Software Tools, Pittsburgh, PA) with one monitor viewed by the experimenter and the other by the participant. The participant sat approximately 50 cm from the monitor and held a microphone connected to the computer's serial port through Eprime's SRbox. Sound input through the microphone provided the stop-signal to a software clock accurate to within ±1 ms.

We used the same problem sets as Campbell and Phenix (2009). Set A was composed of the operands 2, 5, 7, 8 and Set B was composed of the operands 3, 4, 6, 9. Set A included 2 + 5 = 7, 7 + 2 = 9, 8 + 2 = 10, 5 + 5 = 10, 5 + 7 =12, 8 + 5 = 13, 7 + 7 = 14, 7 + 8 = 15 and Set B included 3 + 4 = 7, 6 + 3 = 9, 4 + 4 = 8, 4 + 6 = 10, 9 + 3 = 12, 9 + 4 = 13, 6 + 6 = 12, 6 + 9 = 15. Corresponding multiplication problems were created by substituting a times sign for the plus sign. Eight filler problems composed of one operand from each set were included in the addition pretests and addition and multiplication posttests for a total of 24 problems, however, like Campbell and Phenix, they were not included in the analyses. Problems spanned five characters spaces (e.g., 2 × 3) and were presented in white, Courier New 18 point font against a black background.

Procedure

The experiment took place in a quiet room and required approximately 45 minutes to complete. Participants first received two blocks of the 24 addition problems with instructions to respond quickly and accurately. In the multiplication practice phase, participants either retrieved or studied 40 blocks of the eight MP problems. Half the participants in each group practiced Set A and the other practiced Set B. Participants in the retrieval group were presented with the problem operands (e.g., 5 × 7) and instructed to retrieve the answer to each problem and say it out loud. Participants in the study group were presented both the operands and the correct product (e.g., 5 × 7 = 35) and instructed to silently read each equation and say the displayed answer out loud. Both groups were told they would be tested on the practiced problems later on in the experiment.

For the post-practice test phase participants were instructed to retrieve and state aloud the correct answer. All participants completed two posttest blocks of the 24 addition problems. The addition blocks were separated by a block of multiplication that included the 24 multiplication counterparts. Thus, between the two addition blocks both MP and MU multiplication problems received one retrieval-practice trial.

Each trial began with a one-second central fixation dot and then the problem appeared in horizontal orientation centered at fixation. Response timing began with problem onset and was terminated when the participant's verbal response stopped the timer. The stimulus was instantly removed from the screen, which allowed the experimenter to flag trials on which the microphone failed to detect response onset and mark them as spoiled. The experimenter entered the participant's answer and then the fixation dot appeared to signal the start of the next trial. There was no feedback provided regarding accuracy or speed.

Results

A total of 654 RTs (2.7%) were marked as spoiled by the experimenter and discarded. We report the RT results based on participants' median RT in each analysis cell, but the same pattern of results held for mean RTs

Multiplication Practice

RT.

Mean median correct RT of the multiplication practice phase was entered into a 2 (Group; Retrieval/Study) × 2 (Practice; Early/ Late) ANOVA. Early vs. Late practice was defined as the first three vs. last three multiplication practice blocks. The corresponding means appear in Table 3-1. Participants were slower in early practice (795 ms) compared to late practice blocks (658 ms) [$F(1, 58) = 58.245$, $MSE = 9654.258$, $p < .001$, $\eta p^2 = .501$]. As would be expected, average RT was slower for retrieval (863 ms) than study (590 ms), because for the latter participants simply read aloud the presented product [$F(1, 58) = 60.020$, $MSE = 37115.955$, $p < .001$, $\eta p^2 = .509$]. There was no evidence for a Group × Practice interaction [$F(1, 58) = 0.894$, $MSE = 9655.456$, $p = .348$, $\eta p^2 = .015$]; thus, speed up from early to late in practice was similar with multiplication retrieval (-154 ms) and study (-120 ms).

Table 3-1

Experiment 2 Group by Set Mean Median Response Times (RT's in Milliseconds) and Error Rates (%E) in the Multiplication Practice Phase

	Retrieval		Study	
Problem Set	RT	%E	RT	%E
Early	940 (39.4)	6.6 (1.6)	650 (23.7)	0.0 (0.0)
Late	786 (27.9)	3.6 (1.2)	530 (15.1)	0.1 (0.1)

Percent Error.

The same analysis was conducted on percentage of errors (see Table 2-8). Participants made more multiplication errors early in practice (3.3%), compared to late in practice (1.9%) [$F(1, 58) = 6.193$, $MSE = 10.302$, $p < .016$, $\eta p^2 = .096$]. Retrieval was more error prone (5.1%) than study (0.1%) [$F(1, 58) = 15.977$, $MSE = 48.257$, $p < .001$, $\eta p^2 = .216$]. There was a Practice × Group interaction [$F(1, 58) = 7.429$, $MSE = 10.302$, $p = .008$, $\eta p^2 = .114$]. The retrieval group reduced errors (3.1%) from early to late practice, whereas the error rate for the study group was approximately equal early and late in practice (-0.1%).

Multiplication Posttest

To determine if the multiplication practice trials were similarly beneficial to for the study and retrieval groups, we analyzed the posttest multiplication block, which included both the MP and MU multiplication problems. Equal benefits (i.e., for MP relative MU problems) would eliminate group differences in addition RIF potentially owing to differences in strengthening the practiced multiplication facts, rather than to the type of practice (i.e., retrieval vs. study).

RT.

Mean median correct RT for the multiplication posttest was entered into a 2 (Group; Retrieval/Study) × 2 (MpMu) ANOVA. The means appear in Table 3-2. MP problems (972 ms) were solved faster than MU problems (1216 ms) [$F(1, 58) = 26.998$, $MSE = 66590.848$, $p < .001$, $\eta p^2 = .318$]. These posttest means were comparable to those observed by Campbell and Phenix (2009, Table 1, p. 68) for both MP (948 ms) and MU (1178 ms). There was also a Group × MpMu interaction whereby participants in the retrieval group solved MP problems faster than MU problems (-346 ms, $SE = 61.8$), but the difference for the study group was not as large (-144 ms, $SE = 71.2$) [$F(1, 58) = 4.617$, $MSE = 66590.848$, $p = .036$, $\eta p^2 = .074$]. Thus, unlike Campbell and Phenix, retrieval practice lead to greater RT benefits than did study practice. This result qualifies interpretation of the retrieve-study manipulation with respect to RIF, but either group could still potentially present hyper-RIF.

Percent Error.

The same analysis was conducted on mean percentage of error in the multiplication posttest. The corresponding means appear in Table 3-2. Participants made more errors on MU (11.7%) compared to MP problems (7.1%) [$F(1, 58) = 5.574$, $MSE = 113.057$, $p = .022$, $\eta p^2 = .088$]. All other effects were $p < .10$. Campbell and Phenix (2009, Table 1, p. 68) observed

7.6% errors for posttest MU problems and 3.1% errors for MP problems. Thus, posttest multiplication performance was similar in the present study and the Campbell and Phenix (2009) study.

Table 3-2

Experiment 2 Group by Set Median Response Times (RT's in Milliseconds) and Error Rates (%E) in the Multiplication Posttest

	Retrieval		Study	
Problem Set	RT	%E	RT	%E
MP	909 (43.2)	5.4 (1.9)	1035 (60.5)	8.8 (2.3)
MU	1255 (78.9)	10.0 (1.8)	1178 (67.7)	13.3 (2.7)

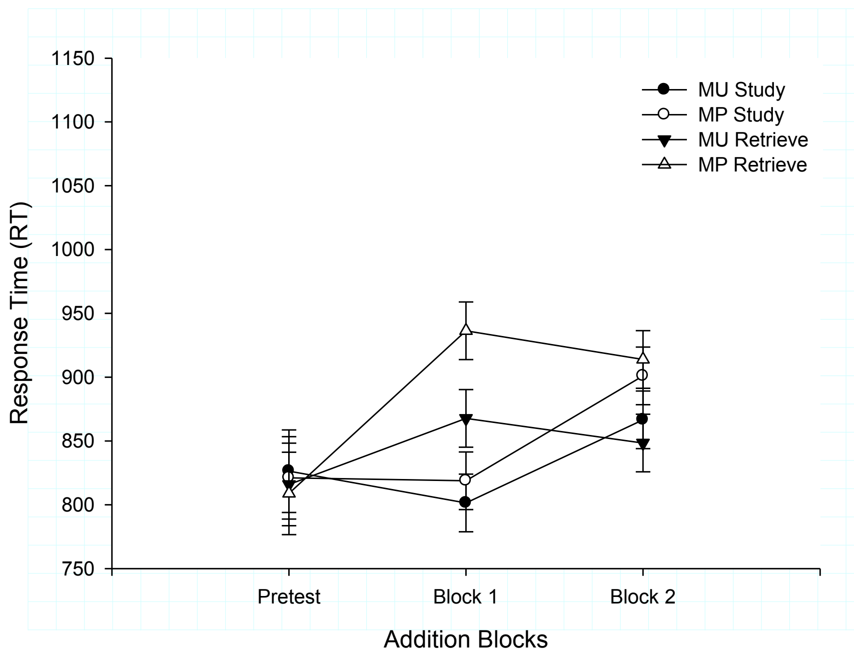
Note. MP = multiplication practiced; MU = multiplication unpracticed.

Addition Pretest

The second (i.e., final) addition pretest block was analyzed to confirm that there were no group differences in addition performance prior to multiplication practice. Mean median correct RT was entered into a 2 (Group; Retrieval/Study) × 2 (MpMu) ANOVA. As Figure 3-1 shows, there were no differences between groups or problem sets in addition RT prior to multiplication practice (all $p < .10$). Similarly, as Figure 3-1 shows, there were no effects of these factors on error rates (all $p < .10$).

Figure 3-1

Experiment 2 Mean Median Correct Addition Response Times (RT) in milliseconds as a Function of Experimental Block, Practice, and Group



Note. Error bars are repeated measure 95% confidence intervals (Jarmasz & Hollands, 2009).

Addition Posttests

RT.

The addition posttests provided the data to assess RIF. A 2 (MpMu) × 2 (Block 1/2) × 2 (Group; Retrieval/Study) ANOVA on mean median RT for correct responses was conducted. The corresponding means appear in Figure 3-1. Overall, participants were slower on MP problems (892 ms) than MU problems (846 ms) [$F(1, 58) = 11.566$, $MSE = 11194.794$, $p = .001$, $\eta p^2 = .166$]. The retrieval group's mean RT increased slightly from Block 1 to Block 2 (+21 ms) but the study group slowed down substantially in Block 2 compared to Block 1 (+74 ms) [$F(1, 58) = 8.296$, $MSE = 16172.948$, $p = .006$, $\eta p^2 = .125$]. The intervening block of multiplication retrieval trials between the addition blocks introduced RIF for the study group in addition Block

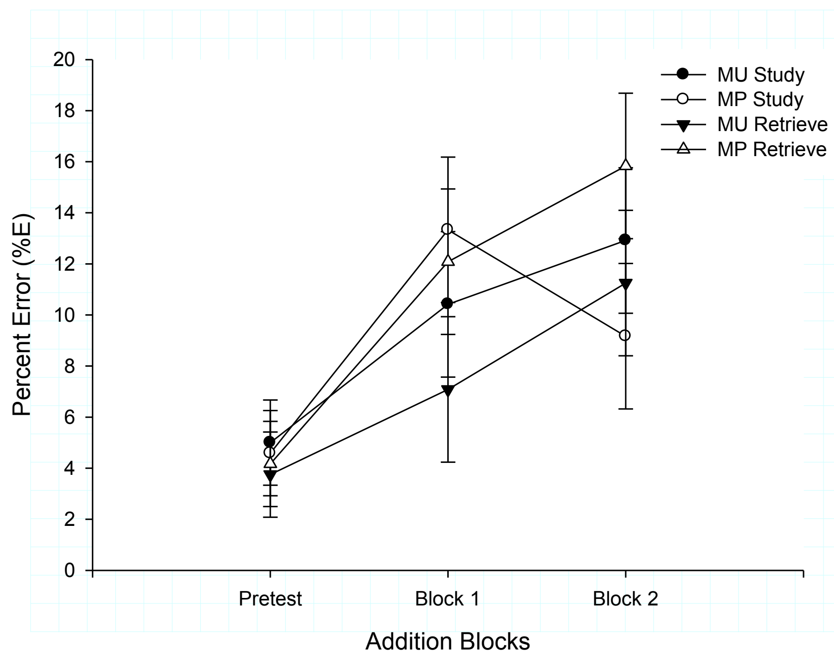
2. Separate MpMu × Block ANOVAs for each group confirmed for the study group a main effect of block only [+74 ms in Block 2, $F\,(1, 29) = 12.819$, $MSE = 12706.250$, $p = .001$, $\eta p^2 = .307$] and for the retrieval group a main effect of MpMu only [a 67 ms RIF effect, $F\,(1, 29) = 9.348$, $MSE = 14427.437$, $p = .005$, $\eta p^2 = .244$]. There were no other significant effects in either group's RT ANOVA. Thus, there was no evidence for hyper-RIF; that is, there was no evidence of greater RIF for MU problems in addition Block 2 after one multiplication retrieval in the posttest, compared to the RIF exhibited by the retrieval group for MP problems in Block 1.

Percent Error.

A corresponding ANOVA was conducted on percent error. The means are in Figure 3-2 to show change across the addition pretest and the two addition posttests. There were no significant effects (all $p \geq .10$).

Figure 3-2

Experiment 2 Mean Percent Error (%E) as a Function of Experimental Block, Practice, and Group



Note. Error bars are repeated measure 95% confidence intervals (Jarmasz & Hollands, 2009)

Discussion

Most important in the present context, neither group presented evidence of the hyper-RIF effect reported by Campbell and Phenix (2009). The hyper-RIF effect would have appeared as much slower RT for MU additions compared to MP additions in Block 2. Instead, for the retrieval group, RT for MP addition problems continued to be slower than MU problems in Block 2. For the study group, both MP and MU were slower in Block 2 than in Block 1, but, if anything, MP problems were answered slower than MU problems. Thus, despite a careful attempt to replicate the conditions that produced hyper-RIF in the Campbell and Phenix study, we were unable to replicate the effect.

In the first addition posttest block, RIF occurred only for the retrieval group (i.e., slower RT for MP additions compared to MU additions) and not the study group. In the second addition posttest block, the retrieval group presented a RIF effect practically equivalent to Block 1. For the study group, both MP and MU problems were slower in Block 2 than in Block 1, which is consistent with RIF induced by the intervening multiplication retrieval practice involving both MP and MU problems. These results are consistent with previous evidence for retrieval-dependent RIF in arithmetic (Campbell and Phenix, 2009; Campbell & Thompson, 2012). Unlike the previous studies, the analysis of the multiplication posttest block showed that retrieval practice benefitted from subsequent retrieval of an item substantially more than study practice. This differential transfer of practice implies that the difference in addition RIF observed between the retrieval and study groups in addition Block 1 might reflect greater strengthening of the multiplication facts in the retrieval condition, rather than an effect due to retrieval vs. study practice per se. Nonetheless, the study group's increase in RT for both MP and MU problems in Block 2 relative to Block 1, following the intervening block of multiplication retrieval practice of MP and MU problems, reinforces the retrieval dependence of RIF in arithmetic memory.

The present failure to observe hyper-RIF in arithmetic memory aligns with the previous failures of Experiment 1 to replicate the effect. We have reconstructed the original Campbell and Phenix (2009) analysis and could find no errors in the results reported by Campbell and Phenix. Comparison of Figures 3-1 and 3-2 (Experiment 2) show that the Campbell and Phenix groups and the present participants were similar in their addition performance and, as noted previously, they were also similar in multiplication performance. This makes differences in arithmetic skill between participants in the two studies an unlikely explanation.

From the research presented here as well as previous research (e.g., Campbell & Dowd, 2012; Campbell & Thompson, 2012), it is clear that the RIF effect is a robust effect in simple addition facts. RIF in arithmetic has also been found in neuropsychological studies (Galfano et al., 2011). Although RIF is a relatively robust phenomenon found across several different studies, we can conclude that hyper-RIF is not. Hyper-RIF is simply an elusive phenomenon, as evidenced by several failed attempts to reproduce the effect reported in Campbell and Phenix (2009) (i.e., Campbell & Thompson, 2012; Experiments 1 and 2 reported here).

CHAPTER 4

General Discussion

Factors that Impact RIF

The hyper-RIF effect has proven to be difficult to replicate. It is possible that the hyper-RIF observed by Campbell and Phenix (2009) was a Type I error and hyper-RIF does not exist. Another possibility is that there are unique boundary conditions for hyper-RIF that we failed to replicate in Experiments 1 or 2. For example, Campbell and Phenix might have tested a sample with unique arithmetic memory characteristics, or perhaps there were undocumented features of the methodology that happened to be crucial for hyper-RIF. In fact, the standard RIF effect is sensitive to many experimental factors and boundary conditions (Anderson, 2003; Storm & Levy, 2012). Given our inability to produce hyper-RIF, the following sections review known properties and boundary conditions for RIF.

The first is the property of cue independence, (i.e., RIF occurs even if cues change). For example, in the typical retrieval-practice paradigm, when particular exemplars are practiced repeatedly (e.g., Fruit- Orange), memory for related items is impaired both when the practiced category cue is presented (e.g., Banana given the practiced category cue Fruit- Ba____) and when a novel, unpracticed category cue is presented (e.g., Monkey- Ba____). This indicates that RIF does not depend on the practiced cue and argues against cue-related blocking or interference as the explanation of RIF (e.g., Anderson & Spellman, 1995).

A second, well-known property of RIF is retrieval specificity, where RIF only occurs when practiced items are retrieved, but not when they are studied. For example, repeated study practice of exemplars (e.g., Fruit- Orange) will not cause RIF to occur on the subsequent recall test. In contrast, however, repeated retrieval practice (e.g., Fruit- Or____ for the cue Orange) will cause RIF to occur at the final recall test. This is because retrieval practice causes inhibition of competitors (i.e., to retrieve an item from memory, competing memories must be supressed), but this does not occur when items are studied (e.g., Anderson & Spellman, 1995). Evidence for retrieval specificity of RIF in arithmetic was provided by Campbell and Phenix (2009), Campbell and Thompson (2012) and Experiment 2 presented here, because RIF for the addition counterparts occurred with retrieval practice of multiplication, but not study.

A third important property of RIF is interference dependence, which suggests that RIF is caused by competition among related items in memory. Retrieval of items and competition must occur for RIF to occur. For example, Anderson et al. (1994) found that RIF only occurred for items that were strongly associated with their cue. For example, RIF occurred with retrieval practice of items that were highly associated with the cue (e.g., Fruit- Orange), but this did not occur for items that were less commonly associated with that cue (e.g., Fruit- Guava). Evidence for interference dependence can also be found in Experiments 1 and 2, because RIF only occurred when there was competition among related items in memory (i.e., competition for those items that have high memory strength). For example, in Experiment 1, when RIF was found, RIF only occurred for the small problems presumably because participants had high memory strength for those problems, so more competition occurred, which caused RIF to occur.

The fourth property of RIF that has been identified is strength independence. Strength independence suggests that the amount of RIF observed is not dependent on the amount that the practiced facts are strengthened. For example, having more practice on some items (e.g., Fruit-Guava) will not cause these items to be more susceptible to RIF compared to other less practiced items (e.g., Fruit—Kiwi). Evidence for strength independence can be seen in Experiment 1, because the three groups that demonstrated an equivalent level of RIF, but the groups differed in terms of the amount of multiplication practice they received in the practice phase (i.e., OpsC/20 vs. OpsC/6 and OpsU/6).

A boundary condition of RIF is integration (e.g., Goodmon & Anderson, 2011). Integration refers to the extent that the associates of a retrieval cue are related by established semantic associations. Specifically, how much the items in the practice phase and the memories that are to be retrieved at final test are integrated, changes the amount of RIF that is observed. The more items are integrated, the less it is that RIF will occur. For example, Goodmon and Anderson (2011) demonstrated that when Rp+ (e.g., Horse) and Nrp (e.g., Pony) were semantically related, then less RIF occurred. Specifically, participants practiced retrieval of horse (Rp+), but the baseline not practiced item was pony (Nrp), which is semantically related to horse and therefore less RIF is observed because the Nrp baseline is reduced (i.e., primed).

The level of RIF that is observed is impacted by the amount of similarity between an associate and a cue. For example, Smith and Hunt (2000) conducted a study where half the participants were asked, during the retrieval practice phase, to find a similarity among the items that they practiced. The other participants were asked to think of differences between the items during the practice phase. The results suggested that when participants were asked to look for similarity among items RIF did occur, but when participants were asked to look for differences among the items, no evidence of RIF was found. For RIF to occur it is important to have strong relations between the categories and their exemplars.

Finally, the amount of RIF that is observed depends on the amount of attention given to the items during the retrieval practice phase. The greater the effort to resolve interference from competing items that occurs, the more likely RIF will occur. For example, Anderson et al. (2000) demonstrated that RIF occurred when participants were asked to complete the category cue on the final recall test (e.g., Fruit- Or____ for the exemplar orange), but not when participants were asked to retrieve the category on the final test (e.g., F____- Orange). Anderson argued that no RIF occurred in the latter condition because the exemplar orange was associated with the category fruit, but not with the other fruit exemplars and therefore no competition was created for these questions.

The known boundary conditions for RIF in arithmetic are retrieval dependency and interference dependence. The other factors reviewed above have not been investigated with arithmetic stimuli. This review of factors and boundary conditions for RIF indicate that it is a highly complex phenomenon that requires a precise balancing of factors to emerge in experimental data. Given this, we cannot rule out the possibility that the Campbell and Phenix (2009) experiment created a unique niche that satisfied all the boundary conditions required by hyper-RIF.

Replication in Psychology

The failure in the present experiments to replicate Campbell and Phenix's (2009) hyper-RIF resonates with growing concerns over problems to replicate published experimental findings in the psychological literature. Replication failure may be promoted because of the difficulty to publish null findings (e.g., the original authors did not report or attempt to publish their own failures to replicate) and by researchers naively or intentionally exaggerating effect sizes. Simmons, Nelson, and Simonsohn (2011) report six requirements for researchers in psychology that would reduce publication bias and promote integrity among researchers. First, when conducting research, authors should determine prior to data collection, when data collection will stop, and report this in the publication. Second, there should be at least 20 participants in each cell of the design (or provide good justification for why this is not possible) (e.g., population, cost). Third, when data is collected, all the variables that are measured and manipulated should be reported, and not just those that yield significant results. Fourth, researchers should report all the experimental conditions that were implemented in the experiment. Fifth, if any data are excluded (e.g., outliers), results with and without this data should be reported. Finally, when a covariate is used in the analysis, researchers should report the data with and without the covariate. Implementation of these guidelines would promote honesty and comprehensiveness in psychological research.

Makel, Pucker, and Hegarty (2012) conducted a meta-analysis of 100 psychology research journals with the highest impact factors since 1900. Out of the articles in these journals only 1.6% included the word "replication". This indicates that replication has been, and still is, a very low priority in the publication of psychological research. A second analysis was conducted on a set of 500 randomly selected articles that used the word replication. Only 68% (n = 352) of these were actual attempted replications of effects. Makel et al., demonstrated that most reported replications were successful, but that the rate of successful replication was significantly higher when the same authors conducted the original and replication studies (91.7%), compared to other authors (64.6%). Of those researchers who conducted the original and replicated studies, only a very small amount (1.8%) reported their own failure to replicate. Other failures to replicate may not be reported because of their inability to be published in academic journals. Thankfully, the Canadian Journal of Experimental Psychology (Editor Douglas J. K. Mewhort) welcomed the submission of Experiment 2.

Models of Arithmetic Memory and RIF

Campbell and Dowd (2012) outlined a model of addition RIF in this paradigm. They proposed that presentation of a target multiplication problem (e.g., 4 × 5) primes a family of related facts including the addition counterpart (e.g., 4 + 5 = 9). If a related fact is a strong competitor of the target memory, it attracts inhibition during target retrieval practice that can produce a net reduction in its accessibility (i.e., RIF). As discussed previously, such competition dependence appears to be a boundary condition for RIF. In contrast, a related fact that is a weak competitor may be primed, but not inhibited, giving a net increase in it's accessibility. Whether the transfer effect of multiplication practice on the addition counterpart is negative (RIF) or positive (priming) depends on the balance of inhibition and priming that results from multiplication retrieval practice.

RIF of arithmetic facts is not addressed by any of the current models of arithmetic memory. Some models do assume that as an arithmetic problem is solved, multiple facts are activated simultaneously in memory (Ashcraft, 1987; Campbell, 1995; Siegler, 1988). This would create the conditions under which resolution of competition would be required and give rise to RIF. Only Campbell's (1995) network interference model incorporates an active mechanism of inhibition. According to this model, when participants are presented with a problem (e.g., 7 × 4), all problem "nodes" (i.e., representations of the facts in long-term memory) that are associated with that problem become activated. For example, if participants are presented with the problem 7 × 4, all the 7-times and 4-times problems are activated, as well as the addition counterpart (4 + 7 = 11). During retrieval, a unit of activation is added to each node based on its similarity to the presented problem, both in terms of common elements (e.g., common operands) and numerical magnitude. Each node also receives inhibition in each processing cycle that is inversely related to each nodes proportion of the total activation in the network. Thus, greater similarity between the presented problem and nodes in the network causes more activation and less inhibition to occur for these items, which results in similar nodes (i.e., categorical and semantically related nodes) in the network to be highly activated and low similarity nodes to be suppressed. This mechanism gives rise to common error characteristics (i.e., errors are often semantically or categorically related to the correct answer; e.g., 9 × 6 = 36) including operation errors (e.g., 3 + 4 = 12). The network interference model encompasses both addition and multiplication as part of the same arithmetic fact representation network and highlights the importance of activation and retrieval inhibition of components. Nonetheless, the network interference model would predict effects opposite to RIF, because the strongest competitors are the most strongly activated. Specifically, more similarity between related items in memory would cause more activation to occur for these items, not suppression of strong competitors. Furthermore, the network interference model assumes that more interference would occur for larger problems in memory, because more competing problem nodes are activated. However, RIF sometimes shows inhibition only of small addition problems (Campbell & Thompson, 2012), presumably because they are stronger competitors to their multiplication counterparts. Thus, the activation and inhibition mechanisms of the network-interference model are ill equipped to produce the RIF phenomena that has now been observed in many experiments. The RIF results published in the last several years represent a basic challenge to researchers in the area of arithmetical cognition, given that no existing model can account for this basic effect.

Conclusions

The basic phenomenon of RIF of addition by multiplication practice is well-established, but no current theory of mental arithmetic accounts for it. The present experiments reaffirm the robustness of the phenomenon and add to the list of replications of the basic effect. The results reported in this thesis, and the related published research on RIF of addition memory, represent an important test for models or cognitive arithmetic. In contrast, the hyper-RIF pattern reported by Campbell and Phenix (2009) is an elusive or non-existent phenomenon; consequently, it cannot be considered an important result in the RIF literature. As such, unless it is subsequently verified by other researchers, it should be given no significant consideration in theoretical models of RIF.

REFERENCES

Abel, M., & Baüml, K-H., T. (2012). Retrieval-induced forgetting, delay and sleep. *Memory, 20*, 420-428. doi: 10.1080/09658211.2012.671832

Anderson, M.C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. Journal of Memory and Language*, 49,* 415-445. doi: 10.1016/j.jml.2003.08.006

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1063-1087. doi: 10.1037/0278-7393.20.5.1063

Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bulletin & Review*, *7*, 522-530. doi: 10.3758/BF03214366

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102,* 68–100.

Ashcraft, M. H. (1987). Children's knowledge of simple arithmetic: A developmental model and simulation. In Brainerd, C.J., Kail, R., Bisanz, J.,(Eds). *Formal Methods in Developmental Research* (pp. 302–338). New York: Springer.

Barnier, A. J., Hung, L., & Conway, M. A. (2004). Retrieval-induced forgetting of emotional and unemotional autobiographical memories. *Cognition and Emotion, 18*, 457-477. doi: 10.1080/0269993034000392

Basden, D. R., & Basden, B. H. (1995). Some tests of the strategy disruption interpretation of part-list cuing inhibition. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* 21,1656-1669.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger and F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309-330). Hillsdale, NJ: Erlbaum.

Blaxton, T. A., & Neely, J. H. (1983). Inhibition from semantically related primes: Evidence of a category-specific inhibition. *Memory & Cognition, 1*, 500-510. doi: 10.3758/BF03196987

Butler, K. M., Williams, C. C., Zachs, R. T., & Maki, R. H. (2001). A limit on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1314-1319. doi: 10.1037/0278-7393.27.5.1314

Camp, G., Pecher, D, & Schmidt, H. G. (2007). No retrieval-induced forgetting using item-specific independent cues: Evidence against a general inhibitory account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 950-958. doi: 10.1037/0278-7393.33.5.950

Campbell, J. I. (1995). Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Mathematical Cognition*, *1*, 121-164.

Campbell, J. I. D., & Dowd, R. R. (2012). Interoperation transfer in Chinese-English bilingual's arithmetic. *Psychonomic Bulletin & Review*. Retrieved June 25, 2012 from http://www.springerlink.com/content/y8084148v0p1r0m7/fulltext.pdf. doi: 10.3758/s13423-012-0277-z

Campbell, J. I. D., & Oliphant, M. (1992). Representation and retrieval of arithmetic facts: A network-interference model and simulation. In J. I. D. Campbell (Ed.), *The nature and origin of mathematical skills* (pp. 331-364). Amsterdam: Elsevier Science.

Campbell, J. I. D., & Phenix, T. L. (2009). Target strength and retrieval-induced forgetting in semantic recall. *Memory & Cognition, 37*, 65-72. doi: 10.3758/MC.37.1.65

Campbell, J. I. D., & Therriault, N. H. (2013, May 17). Retrieval-induced forgetting of arithmetic facts but not rules. *Journal of Cognitive Psychology.* Advance online publication. doi: 10.1080/20445911.2013.798328

Campbell, J. I. D., & Thompson, V. A. (2012). Retrieval-induced forgetting of arithmetic facts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 118-129. doi: 10.1037/a0025056

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1403-1414. doi: 10.1037/0278-7393.25.6.1403

Galfano, G., Penolazzi, B., Fardo, F., Dhooge, E., Angrilli, A., & Umiltà, C. (2011). Neurophysiological markers of retrieval-induced forgetting in multiplication fact retrieval. *Psychophysiology*, *48*, 1681-1691. doi: 10.1111/j.1469-8986.2011.01267.x

Goodmon, L. B., & Anderson, M. C. (2011). Semantic integration as a boundary condition on inhibitory processes in episodic retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition, 37*, 416-436. doi: 10.1037/a0021963

Gómez-Ariza, C. J., Fernandez, A., & Bajo, M. T. (2012). Incidental retrieval-induced forgetting of location information. *Psychonomic Bulletin & Review, 19*, 483-489. doi: 10.3758/s13423-012-0228-8

Hamann, M. S., & Ashcraft, M. H. (1986). Textbook presentations of the basic arithmetic facts. *Cognition and Instruction, 3,* 173-192.

Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review, 11*, 125-130. doi: 10.3758/BF03206471

Jarmasz, J., & Hollands, J. G. (2009). Confidence intervals in repeated-measures designs: The number of observations principle. *Canadian Journal of Experimental Psychology. 63*, 124-138. doi: 10.1037/a0014164

Johnson, S.K., & Anderson, M.C. (2004). The role of inhibitory control in forgetting semantic knowledge. *Psychological Science, 15,* 448-453. doi: 10.1111/j.0956-7976.2004.00700.x

LeFevre, J., Bisanz, J., Daley, K. E., Buffone, L., Greenham, S. L., & Sadesky, G. S. (1996). Multiple routes to solution of single-digit multiplication problems. *Journal of Experimental Psychology: General, 125*, 284-306.

LeFevre, J. A., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology-Learning Memory and Cognition*, *22*, 216-230.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in Psychology Research How *Often* Do They Really Occur?. *Perspectives on Psychological Science*, *7*, 537-542. 10.1177/1745691612460688

Manly, C. F., & Spoehr, K. T. (1999). Mental multiplication: Nothing but the facts? *Memory & Cognition, 27*, 1087-1096. doi: 10.3758/BF03201238

Phenix, T. L., & Campbell, J. I. D. (2004). Effects of multiplication practice on product verification: Integrated structures model or retrieval-induced forgetting? *Memory & Cognition, 32*, 324-335. doi: 10.3758/BF03196862

Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning & Verbal Behavior,* 12, 43-50.

Shaw, J. S., Bjork, R. A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review, 2*, 249-253. doi: 10.3758/BF03210965

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*, 1359-1366. doi: 10.1177/0956797611417632

Shivde, G., & Anderson, M.C. (2001). The role of inhibition in meaning selection: Insights from retrieval-induced forgetting. D. Gorfein (Ed), *On the Consequences of Meaning Selection: Perspectives on Resolving Lexical Ambiguity*, pp. 175-190. Washington, D.C.: American Psychological Association.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General, 117,* 258-275.

Spitzer, B., & Bauml, K. (2007). Retrieval-induced forgetting in item recognition: Evidence for a reduction in general memory strength. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 863-875. doi: 10.1037/0278-7393.33.5.863

Storm, B.C., & Levy, B.J. (2012). A progress report on the inhibitory account of retrieval-induced forgetting. *Memory & Cognition, 40,* 827-843. doi: 10.3758/s13421-012-0211-7

Appendix A

Author's Note

This document serves as Anna Maslany's Masters of Arts thesis in the Department of Psychology at the University of Saskatchewan. Two experiments are reported in the document. Experiment 2 has been published in the Canadian Journal of Experimental Psychology with the supervisor, Dr. Jamie Campbell, as second author. Anna Maslany and Jamie Campbell elected not to publish the results of Experiment 1, however, because the effects of the novel independent variables were equivocal. Experiment 1is presented in detail here, because the results set the stage for Experiment 2 and the ultimate conclusions of this thesis.

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to Jamie Campbell. Address correspondence to Anna Maslany, Department of Psychology, University of Saskatchewan, 9 Campus Drive, Saskatoon, SK, Canada, S7N 5A5 (phone 306-966-6696, fax 306-966-1959, e-mail anna.maslany@usask.ca.