

## Research Reports

# The Obligatory Activation of Practiced Complex Multiplication Facts and What it Tells Us About Models of Arithmetic Processing

Loel Nicholas Tronsky\*<sup>a</sup><sup>[a]</sup> Departments of Psychology and Education, Albertus Magnus College, New Haven, CT, USA.

## Abstract

Three experiments were conducted in which adults practiced complex multiplication problems (e.g.,  $4 \times 17$ ). In Experiments 1 and 2, after practice participants completed a number-matching task in which two digits (cues) were followed by a single digit (probe) and had to determine whether the probe matched either of the cues. In simple arithmetic (e.g.,  $4 \times 3$ ), when the probe is the product of the cues (12), participants are slower/more error prone when determining whether there is a match. Results of Experiment 1 extended this effect to complex multiplication. In Experiment 2, participants practiced problems with the larger operand first (e.g.,  $17 \times 4$ ) or with the smaller operand first (e.g.,  $4 \times 17$ ). The number-matching interference effect from Experiment 1 was replicated, and was equal across the two groups whether cues were presented in their practiced or non-practiced order. Experiment 3 was conducted to determine if two additional simple multiplication effects, consistency and relatedness, could be documented for complex multiplication. After practice, in a verification task ( $4 \times 13 = 56?$ ) it was found that when presented answers shared a digit with the decade digit of the correct answer (consistency) or were a correct answer to another practiced problem (relatedness), participants rejected answers more slowly and/or less accurately. Together, findings from the three experiments support arithmetic models that posit that commuted pairs are not represented in long-term memory independently and that posit representations of two-digit multiplication answers are decomposed into decades and units during arithmetic processing.

**Keywords:** complex multiplication, numerical cognition, arithmetic, number-matching, practice, obligatory activation

Journal of Numerical Cognition, 2016, Vol. 2(2), 140–165, doi:10.5964/jnc.v2i2.22

Received: 2016-01-19. Accepted: 2016-07-11. Published (VoR): 2016-08-05.

\*Corresponding author at: Departments of Psychology and Education, 203 Weldon Hall, Albertus Magnus College, New Haven, CT 06511, USA. E-mail: Ltronsky@albertus.edu



This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

After extensive practice with single-digit arithmetic problems (e.g.,  $8 \times 6$ ), adults mostly change from using deliberate and resource-demanding problem-solving strategies to retrieving answers for those problems directly from long-term memory (e.g., Geary, 1996; Imbo & Vandierendonck, 2008; LeFevre et al., 1996). Many techniques have been used to document arithmetic strategy use; until recently, the most common one has been the use of verbal reports from participants after they solve each problem (e.g., Kirk & Ashcraft, 2001; LeFevre et al., 1996). These studies have helped to uncover the locus of several important effects in basic arithmetic problem solving that have had a profound impact on the development of models of arithmetic processing; an example of one of those effects is the problem size effect, which is evidenced by an increase in error rates and response times as the sums/products of problems increase.

Criticisms of the verbal report technique have been raised, however, calling into question their validity (Kirk & Ashcraft, 2001). As a result, new techniques for determining when problem solving involves retrieval from long-term memory versus strategy use, such as the operand recognition paradigm (Thevenot, Fanget, & Fayol, 2007),

have been developed. For example, [Thevenot et al. \(2007\)](#) had participants solve addition problems in one condition and compare the magnitude of two numbers in another condition. After each trial in both conditions, participants were asked to identify if a third number was one of the original addends from the addition problem or one of the numbers from the comparison task. The argument is that if participants are using strategies to solve problems, the added cognitive resources needed and the interference caused by transitory results (e.g., solving  $5 + 6$  by first arriving at the transitory result of  $5 + 5 = 10$ ) will lead to a decay of the initial problem in memory. Therefore, when strategies are used to solve the arithmetic problem, the number presented after the addition problem should be verified/rejected as matching an addend more slowly than the number presented after the comparison task. This was found to be true for both large and medium problems in [Thevenot et al.'s \(2007\)](#) study suggesting that non-retrieval strategies are used when solving these categories of problems. The validity and promise of the operand recognition paradigm for determining arithmetic solution procedures has been extended to additional operations and samples, including children ([Fanget, Thevenot, Castel, & Fayol, 2015](#); [Thevenot, Castel, Fanget, & Fayol, 2010](#)).

Another technique to determine whether participants are solving problems via retrieval from long-term memory uses tasks for which fact retrieval is irrelevant. For example, [LeFevre, Bisanz, and Mrkonjic \(1988\)](#) had undergraduates engage in a number-matching task. Participants were shown number pairs (e.g., “3” and “4”, called the cue) followed by a single number (the probe). On certain trials the probe was not the sum of the number pair in the cue (e.g., “8”), and on other (sum) trials, the probe was the correct sum of the cue (i.e., “7”). The participants were asked simply to press a button to indicate whether the probe matched either of the cue numbers; therefore, adding the cue numbers was irrelevant to completing the task. What was being tested, however, was whether answers to arithmetic problems were automatically retrieved from long-term memory in arithmetically-skilled participants. If answers were being automatically retrieved, participants’ ability to reject the probe as a match to one of the cues should have been interfered with, leading to slower response times and increased errors. This is exactly what [LeFevre et al. \(1988\)](#) found, and it occurred whether or not an addition sign was included between numbers in the cue.

Since this seminal investigation that supported the obligatory activation of addition facts, subsequent studies have examined the obligatory activation of answers to other operations (e.g., [De Brauwer, 2007](#); [Thibodeau, LeFevre, & Bisanz, 1996](#)), the role of working memory in automatic retrieval of arithmetic facts ([Rusconi, Galfano, Speriani, & Umiltà, 2004](#)), the extension of the obligatory activation to related problems’ answers ( $7 \times 6$  activates 42 and 48; [Galfano, Rusconi, & Umiltà, 2003](#)), and have examined the electrophysiological markers of task-irrelevant retrieval in simple multiplication ([Galfano, Mazza, Angrilli, & Umiltà, 2004](#); [Galfano, Penolazzi, Vervaeck, Angrilli, & Umiltà, 2009](#)).

For example, [Galfano et al. \(2003\)](#) tested whether problem answers plus answers to nearby problems that shared an operand (e.g.,  $7 \times 6$  and  $8 \times 6$ ) were automatically activated in a number-matching task. To do this, they presented cues to participants (e.g., “7” and “6”) followed by probes that were incorrect answers to nearby related-problems’ answers (e.g., “48”—representing the correct answer to  $8 \times 6$ ), and incorrect answers to unrelated problems (“45”—the correct answer to  $9 \times 5$ ). Across the six experiments, participants were significantly slower at rejecting the near-neighbor probes than the unrelated probes, which suggests that the obligatory activation from the operands of a problem extend both to the correct answers and nearby incorrect answers. Results of the number-matching studies outlined above have added to the evidence supporting theories that simple arithmetic problems and answers are stored in associative networks in long-term memory ([Ashcraft, 1992](#); [Campbell, 1995](#)).

When two numbers in a multiplication problem are presented, those numbers are activated, and activation spreads in an obligatory fashion through the network to number nodes that represent the correct product, and to a lesser extent, spread to adjacent nodes that represent correct answers to neighboring problems. It appears that spreading activation in the network is bidirectional, as several experiments have shown that presentation of numbers that are correct answers to multiplication problems automatically activate the operands of those problems (Rusconi, Galfano, Rebonato, & Umiltá, 2006).

While there has been a proliferation of research examining simple arithmetic processing, the research examining complex arithmetic processing is relatively sparse. Much of this research has focused on examining working memory involvement (e.g., Imbo & LeFevre, 2010; Tronsky, 2005), strategies used by typical (skilled adults) and atypical (e.g., math precocious/disabled students, calculating prodigies) populations (e.g., Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Hoard, Geary, Byrd-Craven, & Nugent, 2008; Pesenti, Seron, Samson, & Duroux, 1999), and brain areas recruited during the learning and processing of complex arithmetic problems (e.g., Grabner, Rüttsche, Ruff, & Hauser, 2015; Ischebeck, Zamarian, Egger, Schocke, & Delazer, 2007). To date, there have not been any investigations of complex arithmetic processing using a number-matching task. Therefore, the goal of the first experiment is to establish that the obligatory activation of problem answers extends from simple to complex arithmetic.

## Experiment 1

In this first experiment, participants were asked to practice solving a small subset of complex multiplication problems that had one multi-digit operand between 14 and 19 and one single-digit operand between 3 and 6. After practice, participants completed a number-matching task modeled after that used by Galfano et al. (2004). The hypothesis for Experiment 1 is that after practice, interference in the number-matching task will be demonstrated by slower response times and higher error rates for probes that are products of the cues compared to probes that are unrelated to the cues. This will be indicative of the obligatory activation of the product of the cue numbers, which has been documented in the simple arithmetic experiments outlined above.

### Method

#### Participants

Twenty-one undergraduates (16 female, 5 male) from a small Northeastern college in the USA participated and received extra credit toward their psychology classes; their mean age was 19.4 years. All had normal or corrected-to-normal vision. Three participants did not have English as their first language, and one reported a diagnosed math disability. Patterns of results did not differ when these four participants' data were removed from analyses, so their data were retained in the analyses that are reported.

The computer tasks described below were administered using the SuperLab Pro experiment software (version 4.5) with responses made either verbally into a microphone (complex multiplication practice) or via a response-box button press (number-matching task). Stimuli in all tasks were displayed on a computer monitor in black Tahoma Regular 36-point font against a white background. Double-digit numbers were a width of approximately 12 mm and single-digit numbers a width of approximately 6 mm; both were 10 mm in height. Participants viewed these stimuli at a distance of approximately 60 centimeters from the computer monitor.

**Complex multiplication practice stimuli** — Five complex multiplication problems ( $16 \times 3$ ,  $18 \times 3$ ,  $17 \times 4$ ,  $19 \times 4$ , and  $14 \times 6$ ) and their commuted pairs were presented on a computer screen for participants to solve mentally. In total, the multiplication problems covered a width of approximately 4.5 cm on the screen. These problems were presented in blocks of ten (each problem and its commuted pair), and within each of the blocks, problems were presented in a random order.

**Number matching stimuli** — As noted above, the stimuli for this task were (mostly) modeled after those that have been used in number matching tasks examining the automatic activation of simple multiplication answers (e.g., Galfano et al., 2004; Rusconi et al., 2004). Recall that the number-matching task involves the presentation of two cue numbers followed by a probe number, and participants determine whether the probe number matches one of the cue numbers. As in recent investigations using this task, the cue numbers were presented without a multiplication sign between them (e.g., Galfano et al., 2009). They were separated by three spaces, which made them approximately 3 cm wide. Half of the stimuli included a probe number that did not match either of the cue numbers, and half of the stimuli included a probe that matched one of the cues. The specific stimuli for Experiment 1 are in Appendix A and are described below.

There were two main categories of stimuli constructed for this task, non-matching stimuli and matching stimuli. Within the non-matching stimuli were three categories of stimuli including product, unrelated, and filler stimuli. The non-matching product stimuli included probes that were the correct answers to the product of the cues (e.g., cues: “16” and “3”; probe: “48”). Non-matching unrelated stimuli included probes that were not the correct answers to the product of the cues (e.g., cues: “16” and “3”; probe: “52”). The non-matching filler stimuli included double-digit and single-digit cues followed by a non-matching double-digit probe (e.g., cues: “67” and “8”; probe: “58”). These items were included so that non-matching item cues were not only the operands of the practiced problems and so there were items with double-digit cues beyond the teen numbers.

The non-matching unrelated items were constructed using several criteria to ensure that the hypothesis, interference from product probes stems from automatic activation of correct answers by the cues, could be tested without confounds stemming from the nature of the probe items.

First, the odd-even (parity) status between probes in the product and unrelated trials were matched—in fact, all probes were even numbers; previous research has shown that different processing strategies may be used during the processing of arithmetic verification problems when parity of the proposed answer does not match the parity of the correct answer (e.g., Lemaire & Fayol, 1995). Next, each probe in the product and neutral conditions appeared the same number of times. Third, most of the unrelated probes were not multiples of either of the cue numbers in a trial with the exception of “72” for the “14” and “6” cue and “92” for the “19” and “4” cue. Even though both of these probes are multiples of the single digit number in each trial, minimal interference, if any, should occur given that it is unlikely that participants would have strong activation of “72” or “92;” it is improbable that participants had the corresponding problems ( $6 \times 12$  and  $23 \times 4$ ) strongly associated with the aforementioned probes. Also, the average magnitude of the probes in the production trials (66) was very close to the average magnitude of the probes in the unrelated trials (64); this is important because it rules out the possibility that participants would respond more slowly to product trials because the magnitude differences between cues and probes were smaller than for unrelated trials (e.g., Dehaene & Akhavan, 1995; van Opstal & Verguts, 2011). Finally, the number of partial matches between cues and probes (e.g., cue: “18” and “3”; probe: “58”; the ones digit matches) was minimized to reduce the potential interference from that structural variable—only the product trial with a cue of “14” and “6”

and a probe of “84” involved a partial match (although there is evidence that partial matches may not influence response time data in the context of this task—see [Galfano et al., 2003](#))

The other category of items in the number-matching task was the matching stimuli. These stimuli also included three sub-categories: probe-balancing stimuli, cue-balancing stimuli, and filler stimuli. The matching probe-balancing stimuli included the same probes as the non-matching product stimuli, but all double-digit probes matched the double-digit number in the cue (e.g., cues: “48” and “3”; probe: “48”). Matching cue-balancing stimuli were when the same cues as the non-matching product trials were used, and cues were followed by probes that matched the single-digit number in the cue (e.g., cues: “16” and “3”; probe: “3”). Lastly, the matching-filler stimuli were included to balance the number of matching and non-matching trials; each contained a double-digit number and single digit number as a cue, and the probes matched the double-digit number in the cue (e.g., cues: “67” and “8”; probe: “67”).

In total, 60 trials of stimuli (cues and probe), ten in each of the six different categories described above, were included. Half of the items in each category had cues with the larger number on the left, and half had cues with the larger number on the right. Fourteen blocks with 30 trials in each block were administered for a total of 420 trials; within each block, trials were presented in a random order, half of the trials were matching and half were non-matching, and half of the trials from each of the categories were those with the larger cue number on the left.

### Procedure

Participants signed up for two separate appointments that were not to be more than one week apart. During the first appointment, participants practiced solving the complex multiplication problems for 60 minutes, and during the second appointment they practiced the problems for an additional 15 minutes before completing the number matching task that took approximately 25 to 30 minutes.

Before data collection began for the complex multiplication practice task, a research assistant read a set of instructions that described the problem-solving strategy participants were to use. This procedure involved multiplying the single digit operand by the tens digit of the double-digit operand, multiplying the single digit by the ones digit of the double-digit operand, and then adding together the partial products to arrive at an answer. For example, to solve the problem  $18 \times 3$ , participants would use the following steps:  $10 \times 3 = 30$ ;  $3 \times 8 = 24$ ;  $30 + 24 = 54$ ; the name “tens strategy” was used to identify this procedure. After the instructions, participants completed a set of five warm-up problems that were different from the practice problems so that they could become accustomed to the use of the tens strategy, could become familiar with voicing answers into the microphone, and could ask questions about the procedure. During the data collection phase of the task, the multiplication problems remained on the screen until the microphone was tripped by participants’ vocalization of an answer. Once the problem had disappeared, the word “STRATEGY” was displayed for 1500 milliseconds (ms); participants responded “Yes” if they had used the “tens strategy,” or “No” if they had simply remembered the answer (retrieved the answer from long-term memory). A research assistant recorded participants’ answers and strategy use for each problem. Once the 1500 ms time limit had expired, a white screen was presented for 1000 ms followed by the next practice problem. After completing two blocks of problems (20 in total), participants were offered the opportunity to take a short break to help reduce visual and attentional fatigue. Once 60 minutes had elapsed, participants were released from the session.

As noted, during the first 15 minutes of their second appointment, participants solved the practice problems before completing the number-matching task. A research assistant read instructions that explained the nature of the number-matching task and instructed participants to respond as quickly and accurately as possible. Responses were made using the index fingers of each hand with roughly half of the participants' randomly assigned to use the left key to indicate the probe matched one of the cues, and the other half to use the right key to indicate a match. A set of 35 warm-up trials (different from those used in the data collection portion of the task) were completed so participants could get accustomed to the speed of presentation of stimuli and become familiar with the appropriate key to press for matching versus non-matching trials.

The sequence and timing of events during number-matching trials was as follows (see [Appendix B](#) for a visual representation of this). Each trial began with the presentation of a fixation stimulus (#) in the middle of the screen for 400 ms to orient the participant. This was followed by the presentation of the cue numbers (e.g., "18" and "3") for 80 ms and then four pound signs (####) for 40 ms as a masking stimulus. Next, for 140 ms a white screen appeared as an inter-stimulus interval (ISI). Thus, the stimulus onset asynchrony (SOA) was 260 ms, which is similar to the moderate-length SOAs that have been used in other experiments involving the number-matching task ([Galfano et al., 2003, 2004](#)). Due to the short practice duration, activation of complex problem answers most likely will be slower compared to activation of simple problem answers that have been studied previously; using the moderate-length SOA will make it more likely that potential interference effects will be captured. Finally, a probe number (e.g., 54) was presented for 2500 ms or until participants made a response. After participants had determined whether the probe number matched either of the cue numbers, they pressed the appropriate button on the response pad, and then another white screen appeared for 1000 ms before the next trial began. Stimuli were presented in a unique random order for each participant. Upon completion of the number-matching task, participants were debriefed, allowed to ask any questions, and were thanked for their participation before being released from the experiment.

## Results and Discussion

It should be noted that conventional null-hypothesis significance testing (NHST) was conducted throughout the manuscript; however, since many researchers have noted the shortcomings of NHST, Bayesian analyses also were conducted (using the freeware *MorePower 6.0* for ANOVA; [Campbell & Thompson, 2012](#)) and the resulting Bayes factors (*BFs*) reported. A lengthy exposition of the benefits of using Bayesian analyses is beyond the scope of this article, but a core benefit is that it enables one to determine the relative likelihood that the data fit the alternative hypothesis or the null hypothesis using the observed data, rather than simply rejecting or accepting the null hypothesis using hypothetical data (the sampling distribution) as is done in NHST. For comprehensive explanations of the benefits of Bayesian analyses and how to conduct them one can consult [Jarosz & Wiley \(2014\)](#); [Masson \(2011\)](#); [Wagenmakers \(2007\)](#); and/or [Wagenmakers, Morey, & Lee \(2016\)](#).

To document the problem-solving skill development over the course of the practice sessions in all three experiments median RT data, accuracy rates, and reported use of retrieval was averaged across the first 20, middle 20, and final 20 problems participants solved; please refer to [Table 1](#) for a summary of these data. Experimenter error and equipment failure (computer and microphone) led to some loss of information in each experiment, but all sets of data are based on at least 70% of the participants in each sample. A series of one-factor ANOVAs was conducted on the RT, accuracy, and use of retrieval data from all three experiments. These analyses support the significant skill development of each sample; across practice, latencies decreased significantly,  $F_s > 30.36$ ,  $p < .001$ ,  $\eta^2_p s > .56$ ,  $BF_{10} s > 100$ , while proportion of problems solved correctly increased,  $F_s > 4.68$ ,  $p < .05$ ,  $\eta^2_p s > .20$ ,  $BF_{10} s$

> 2.06, as did reported use of retrieval,  $F_s > 6.12$ ,  $p < .01$ ,  $\eta^2_p > .42$ ,  $BF_{10} > 8.09$ . The average final problem-solving latencies across the experiments (2760, 2239, and 2380 ms) were in line with previous complex multiplication practice investigations (Grabner et al., 2009; Ischebeck, Zamarian, Schocke, & Delazer, 2009).

Turning to the number-matching task, only product and unrelated trials were analyzed, as they were the critical trials needed to test the hypothesis that product probes would automatically activate the correct product of the cues; this analysis procedure follows previous investigations of this type (e.g., Galfano et al., 2009; LeFevre et al., 1988; Rusconi et al., 2006; Thibodeau et al., 1996). In .2% of the trials participants did not make a response before the 2500 ms cut-off (equal for both the product and unrelated trials). Only correct response RTs were analyzed, and to reduce the influence of outliers, data trimming procedures developed by Van Selst and Jolicoeur (1994) used in a previous number-matching experiment (Galfano et al., 2004) were employed within participants, separately for each probe-type condition. These procedures were implemented using routines in the statistical package R (R Development Core Team, 2012) and involve the use of a moving criterion for determining the standard deviations used to trim the data. That is, the standard deviations used to calculate cut-off points are determined by the size of the sample of RTs in each condition and therefore are not biased, as they can be when researchers (somewhat arbitrarily) choose them. Three different methods were offered by Van Selst and Jolicoeur—modified recursive, non-recursive, and hybrid approaches (see their article for an in-depth explanation of each). Given that the modified recursive and non-recursive methods tend to result in divergent trends, the RT analyses across the three experiments were run after using the hybrid trimming method, as it is an average of the other two methods. Analyses were conducted after employing the recursive and non-recursive trimming methods as well, but results of those are mentioned only when they conflict with the hybrid-related findings.

The hybrid trimming procedure resulted in the removal of 3.3% of the product probe and 3.0% of the unrelated probe RTs. Data then were averaged by participant by probe type and submitted to a one-tailed correlated  $t$ -test. The analysis indicated that the 19 ms slower response time associated with the product probes ( $M = 532$  ms;  $SD = 101$ ) versus unrelated probes ( $M = 513$  ms;  $SD = 100.0$ ) was significant,  $t(20) = 3.54$ ,  $p < .001$ , Cohen's  $d = .19$ ,  $BF_{10} = 36.1$ . The Bayes factor indicates that the data were 36.1 times more likely to occur under the model including an effect of probe type than one without it.

A second one-tailed correlated  $t$ -test was run on the accuracy data. It revealed that the proportion of correct responses for the product probes ( $M = .98$ ;  $SD = .03$ ) was significantly smaller than for the unrelated probes ( $M = .99$ ;  $SD = .02$ ),  $t(20) = 2.48$ ,  $p < .05$ , Cohen's  $d = .37$ ,  $BF_{10} = 1.63$ . Review of the Q-Q plots from the analysis showed the residuals deviated from a normal distribution. As such a non-parametric test, the Wilcoxon signed-ranks test, was performed and confirmed that participants' accuracy for the product probe trials was significantly poorer than for the unrelated probe trials,  $Z = 2.02$ ,  $p < .05$ .

Finally, recall that two cues in the number-matching task ("14" and "6") and ("6" and "14") occurred with a product probe ("84") that had a partial match to one of the cue numbers—the number "4" in the ones place of the two-digit cue and in the two-digit probe, match. Both RT and accuracy analyses were re-run after excluding RT and error data related to trials with the aforementioned cues and after using the hybrid process to trim RTs. The significant effects from the initial analyses were unchanged; participants completed product-probe trials 21 ms more slowly than unrelated-probe trials,  $t(20) = 3.81$ ,  $p < .05$ , Cohen's  $d = .20$ ,  $BF_{10} = 67.2$ , and 1% less accurately,  $t(20) = 1.82$ ,  $p < .05$ , Cohen's  $d = .37$ ,  $BF_{10} = 1.09$ ;  $Z = 1.76$ ,  $p < .05$  (Wilcoxon signed-ranks test). These results

Table 1

Mean Practice Task Response Times, Accuracies, and Self-Reported Use of Retrieval

	Initial 20 Problems	Middle 20 Problems	Final 20 Problems
<b>Response Time</b>			
Experiment 1	8488 (2977)	4214 (1837)	2760 (1195)
Experiment 2	8597 (3753)	3057 (1940)	2239 (1137)
Experiment 3	8075 (4957)	3092 (1310)	2380 (978.6)
<b>Accuracy</b>			
Experiment 1	.75 (.25)	.88 (.13)	.94 (.12)
Experiment 2	.77 (.24)	.93 (.11)	.96 (.08)
Experiment 3	.83 (.29)	.97 (.06)	.99 (.02)
<b>Retrieval Use</b>			
Experiment 1	.16 (.33)	.51 (.41)	.59 (.45)
Experiment 2	.11 (.19)	.71 (.37)	.81 (.31)
Experiment 3	.22 (.35)	.78 (.38)	.94 (.20)

Note. Response times are in milliseconds and accuracies are proportion correct. Standard deviations are in parentheses.

lend support to the findings of [Galfano et al. \(2003\)](#) that partial matches in the number-matching task most likely do not contribute to the interference effect.

In summary, the results of the first experiment show that after a little over an hour of practicing complex multiplication problems, participants demonstrated an interference effect during a number-matching task that suggests retrieving the answers to the complex multiplication problems was obligatory. It is especially compelling because multiplication of the cue numbers is irrelevant to what participants were instructed to do in the task. This finding mirrors both the nature and size of the interference effects that have been documented in recent investigations of simple multiplication. The effects in the present experiment were a 19 ms slowing and a 1% error rate increase, and are similar to those found in other studies, which have ranged from 18 to 67 ms (many have been between 18 and 25 ms) and from 2% to 4% errors ([Galfano et al., 2004](#); [Galfano et al., 2009](#); [Rusconi et al., 2004](#); [Thibodeau et al., 1996](#)). Comparing the effect sizes (Cohen's *d*) yields similar results across these studies as well—.19 in the present study compared to .14 to .26 (my calculations) in previous experiments.

## Experiment 2

Given the findings from Experiment 1 that parallel effects from simple multiplication, the purpose of the second experiment was twofold. First, it was to replicate the number-matching interference effect from Experiment 1 with a new sample. Second, it was to use complex multiplication practice and the number-matching task to investigate/evaluate existing models of arithmetic. One model of how multiplication facts are stored in long-term memory is the Identical Elements (IE) Model ([Rickard, 2005](#)). In this model, number triplets (e.g., 3, 8, 24) exist for multiplication and division, and these triplets are stored as three distinct units:  $3 \times 8 \leftrightarrow 24$ ,  $24 \div 8 \rightarrow 3$ , and  $24 \div 3 \rightarrow 8$ . Because the present investigation involves multiplication, the first set of triplets is the most relevant. Rickard claims that multiplication problems and their commuted pairs are not stored in long-term memory as separate entities but instead as one unit. In other words,  $8 \times 3 = 24$  and  $3 \times 8 = 24$  are stored in the same representation. Also, at least for multiplication, there is bidirectional activation that spreads to the answer when the operands are

presented and that spreads to operands when an answer is presented (24 activates  $3 \times 8/8 \times 3$  and  $4 \times 6/6 \times 4$ ; Rickard, 2005; Rusconi et al., 2006).

Some predictions of the IE model related to multiplication problem-answer representations are that RT changes from practicing a particular operand order should transfer to the reverse operand order, and transfer should result regardless of the format in which problems are practiced (e.g.,  $8 \times 3$ , eight x three, or auditory presentation of  $8 \times 3$ ). For example, Rickard, Healy, and Bourne (1994) had participants extensively practice multiplication (e.g.,  $\_\_\_ = 9 \times 6$ ) problems in one operand order and then after practice tested them on both the practiced and non-practiced operand orders. They found that participants demonstrated significantly faster RTs as a result of practice for the non-practiced problem order, but these were still 80 ms slower than for the practiced order. Subsequent experiments replicated this transfer effect and demonstrated that the RT difference between operand orders after practice was likely due to perceptual factors—visual processing of the practiced operand order led to a perceptual speed advantage over the commuted pair at post-practice (Rickard & Bourne, 1996). Neuropsychological evidence also supports the IE model; after brain damage, participants' impaired/spared multiplication abilities are strongly correlated across commuted pairs—if a participant cannot retrieve the answer to  $8 \times 3$ , most likely s/he will not be able to retrieve the answer to  $3 \times 8$  (Hittmair-Delazer, Semenza, & Denes, 1994; McCloskey, Aliminosa, & Sokol, 1991).

In contrast, there are researchers who propose that multiplication problems' commuted pairs are represented independently. In these models, problems and answers are typically characterized as being part of a network with the physical format of the problem ( $8 \times 3$  vs. eight x three) influencing not just perceptual processes, but central cognitive processes as well. For example, in Campbell's encoding complex model (Campbell, 1994, 1995), the physical format of the problem is posited to have an impact on calculation processes (among others), such as the probability of using memory retrieval or strategy-based procedures to arrive at an answer (e.g., Campbell & Albers, 2009). Most important for the present discussion and the rationale for Experiment 2 is that in these models, problems are represented independently both in larger operand first and smaller operand first forms (e.g., Ashcraft, 1992; Campbell, 1995).

To evaluate further the models of arithmetic processing just described, Experiment 2 used a design that included complex multiplication practice and a number-matching task. Participants again were tasked with practicing a small subset of complex multiplication problems before completing a related number-matching task. The one difference from Experiment 1 to Experiment 2 was the use of two practice conditions. Participants in one condition practiced only complex multiplication problems that had the larger operand first (e.g.,  $17 \times 4$ — $L \times s$  problems), and in the other condition participants practiced the same problems that had the smaller operand first (e.g.,  $4 \times 17$ — $s \times L$  problems). In the subsequent number-matching task, participants were shown cues in both orders (" $17$ " and " $4$ ", and " $4$ " and " $17$ "). The different arithmetic representation models outlined above lead to different predictions about the interference that will result in the number-matching task. The IE model predicts that in both practice conditions, a single problem-answer representation will be strengthened for each practiced problem. Therefore, the interference effect observed in Experiment 1 should occur for participants in both practice conditions equally on the  $L \times s$  and  $s \times L$  probes in the number-matching task. Independent operand representation models (e.g., Campbell, 1995) predict that there will be a three-way interaction of practice condition, presentation order (cue type), and probe type—number-matching task interference will be greater when the order of the numbers in the probe match the operand order that was practiced.

## Method

### Participants

Sixty-four undergraduates from a small Northeastern college in the USA volunteered to participate in the experiment. Four participants did not attend their second appointment leaving a total of 60 participants (47 females and 13 males) with a mean age of 21.7 years (range was 18 to 48 years) who received extra credit toward their psychology classes for completing the experiment. All had normal or corrected-to-normal vision. Five participants did not have English as their first language, and none reported a diagnosed math disability or attention deficit disorder. Patterns of results did not differ when removing the five participants' data, so their data were retained in the analyses below. The apparatus and computer tasks were the same as in Experiment 1, although there were minor differences in stimuli used in Experiment 2, and those are explained below.

**Complex multiplication practice stimuli** — The same complex multiplication problems from Experiment 1 were used with the exception of the problems  $14 \times 6$  and  $6 \times 14$ , which were replaced by the problems  $15 \times 6$  and  $6 \times 15$ . This was done so that none of the non-matching product trials in the number-matching task would have a probe that was a partial match to either of the cue digits. Problems were presented in blocks of five, and within each of the blocks, they were presented in a random order.

**Number matching stimuli** — The stimuli for this task were the same as in Experiment 1 except for the product, probe-balancing, and cue-balancing trials related to the two practice problems that were replaced (see [Appendix C](#) for the full set of stimuli).

### Procedure

Participants signed up for two appointments that were not more than one week apart. After signing the consent form, participants were assigned randomly either to the condition in which practice problems were presented larger operand first ( $L \times s$  condition) or smaller operand first ( $s \times L$  condition). During this first appointment, participants practiced solving the complex multiplication problems for 60 minutes, and during the second appointment practiced the problems for 15 minutes before completing the number-matching task that took approximately 25 minutes. The remaining procedures for the experiment and within the experimental tasks (e.g., sequence of stimuli in number-matching task) were identical to Experiment 1.

## Results and Discussion

Once again, only product and unrelated trials were analyzed because they were the critical trials needed to test the hypotheses. Due to computer issues, 11 participants did not complete the full number-matching task, which would have resulted in 35 trials per condition, but each of these participants had at least 22 trials per condition and therefore remained in the data analysis. Overall, in 1.0% of the trials participants did not make a response before the 2500 ms cut-off—1.2% for the  $s \times L$  product condition, .8% for the  $s \times L$  unrelated condition, .8% for the  $L \times s$  product condition, and 1.1% for the  $L \times s$  unrelated condition. The hybrid trimming procedure was used again resulting in the removal of 3.5% of RTs in the  $s \times L$  product, 2.7% in the  $s \times L$  unrelated, 3.0% in the  $L \times s$  product, and 2.9% in the  $L \times s$  unrelated condition. The means of those trimmed RTs then were submitted to a 2 (Practice Type:  $s \times L$  vs.  $L \times s$ )  $\times$  2 (Cue Type:  $s \times L$  vs.  $L \times s$ )  $\times$  2 (Probe Type: product vs. unrelated) ANOVA with repeated measures on the last two factors. For a summary of the means and standard deviations related to the ANOVA analyses, please see [Table 2](#).

Table 2

Mean Number-Matching Task Response Times and Accuracies by Practice Set and Cue Type

Practice	s x L Product	s x L Unrelated	L x s Product	L x s Unrelated
<b>RTs</b>				
s x L	554 (98.6)	551 (106)	571 (110)	559 (113)
L x s	578 (126)	562 (142)	602 (150)	566 (116)
<b>Accuracies</b>				
s x L	.95 (.11)	.97 (.08)	.95 (.09)	.97 (.09)
L x s	.94 (.10)	.96 (.08)	.94 (.09)	.95 (.09)

Note. Response times are in milliseconds and accuracies are proportion correct. Standard deviations are in parentheses.

The 20 millisecond difference in number-matching RTs between the s x L and L x s practice conditions resulted in a non-significant main effect of practice type,  $F(1, 58) = .34, p > .05, \eta^2_p < .01, BF_{10} = .16$ . There was a main effect of cue type,  $F(1, 58) = 5.30, p < .05, \eta^2_p = .08, BF_{10} = 1.78$ ; s x L cues ( $M = 561; SD = 116$ ) were solved 13 ms faster than L x s cues ( $M = 574; SD = 120$ ). Also, the main effect of probe type was significant,  $F(1, 58) = 11.83, p < .001, \eta^2_p = .17, BF_{10} = 33.9$ , as product-probe trials ( $M = 576; SD = 118$ ) were solved 17 ms slower than unrelated-probe trials ( $M = 559; SD = 116$ ); the Bayes factor indicates that the data were 33.9 times more likely to occur under the model including an effect of probe type than one without it, which was almost the same magnitude for the probe effect in Experiment 1 ( $BF_{10} = 36.1$ ). The two-way interactions were not significant,  $F_s < 3.70, ps > .05, \eta^2_{ps} < .07, BF_{10}s < .82$ , and the three-way interaction was not significant either,  $F(1, 58) = .45, p > .05, \eta^2_p < .01, BF_{10} = .16$ . The last Bayes factor indicates that the data were 6.14 times more likely to occur under a model that does not include a practice x cue type x probe type interaction than one that does. It should be noted that analysis of the data after using the modified recursive trimming procedure yielded an additional significant effect—the probe type x practice interaction,  $F(1, 58) = 4.94, p < .05, \eta^2_p = .08, BF_{10} = 1.51$ . According to the Bayes factor, however, this is considered to be weak/anecdotal evidence for the effect (Jarosz & Wiley, 2014).

Separate follow-up 2 x 2 ANOVAs for each practice group resulted in null effects for cue type for both groups,  $F_s < 3.22, ps > .05, \eta^2_{ps} < .10, BF_{10}s < .34$ . The probe-type effect was significant for the L x s practice group, as they were 26 ms slower responding to the product versus unrelated trials,  $F(1, 26) = 9.27, p < .01, \eta^2_p = .26, BF_{10} = 10.3$ , but it was not significant (8 ms slower) for the s x L practice group,  $F(1, 32) = 1.92, p > .05, \eta^2_p = .06, BF_{10} = .63$ . Neither of the cue x probe type interactions was significant,  $F_s < 1.66, ps > .05, \eta^2_{ps} < .07, BF_{10}s < .30$ .

Due to experimenter error, one participant's accuracy data were lost resulting in one fewer degree of freedom in the analysis below. Accuracy on the number-matching task was, in general, high across participants. The mean accuracy data were submitted to the same 2 x 2 x 2 ANOVA that the RT data were. Neither the main effects of practice nor cue type were significant,  $F_s < .28, ps > .05, \eta^2_{ps} < .01, BF_{10}s < .15$ , but there was a main effect of probe type,  $F(1, 57) = 12.58, p < .01, \eta^2_p = .18, BF_{10} = 9.71$ ; proportion of correct responses for the product probes ( $M = .94; SD = .07$ ) was smaller than for the unrelated probes ( $M = .96; SD = .08$ ). None of the interaction effects were significant,  $F_s < .49, ps > .05, \eta^2_{ps} < .01, BF_{10}s < .19$ . As in Experiment 1, however, the ANOVA residuals were not normally distributed, so multiple Wilcoxon sign-rank tests were also performed to examine the between participant effects and Mann-Whitney U tests to examine the between participant and mixed effects. The results

exactly mirrored the ANOVA findings as the effect of probe type was significant,  $Z = 3.22$ ,  $p < .001$ , and no other main or interaction effects reached significance,  $Z_s < .81$ ,  $p_s > .05$ .

Separate follow-up 2 x 2 ANOVAs for each practice group indicated that a probe-type error effect was found for the L x s practice group (1% more errors for product versus unrelated probes),  $F(1, 31) = 4.80$ ,  $p < .05$ ,  $\eta^2_p = .13$ ,  $BF_{10} = 1.77$ , as well as for the s x L practice group (2% more errors for product versus unrelated probes),  $F(1, 26) = 8.91$ ,  $p < .01$ ,  $\eta^2_p = .26$ ,  $BF_{10} = 10.3$ . There were no significant differences for either group related to the cue type main effect,  $F_s < .04$ ,  $p_s > .05$ ,  $\eta^2_{ps} < .01$ ,  $BF_{10s} < .20$ , or related to the probe type x cue type interaction,  $F_s < 1.28$ ,  $p_s > .05$ ,  $\eta^2_{ps} < .04$ ,  $BF_{10s} < .34$ . Again, Wilcoxon sign-rank tests were performed and matched the ANOVA results; a main effect of probe type was found for both the s x L practice group,  $Z_s = 2.34$ ,  $p_s < .01$ , and the L x s practice group,  $Z = 2.17$ ,  $p < .05$ , and the cue type main effect and interaction effect were not significant for either practice group,  $Z_s < .72$ ,  $p_s > .05$ .

In summary, the first finding of note is that the interference effect shown both in longer RTs (driven by the L x s group) and higher error rates (driven by both practice groups) for product vs. unrelated trials was replicated in Experiment 2. Partial eta squared values suggest that these effects were of moderate size, as they explained 17% of the within subject variance for RTs (in separate group analyses, 26% for L x s and 8% for s x L groups) and 18% of the within subject variance for accuracy (13% for L x s and 26% for s x L groups). In addition, results of the Bayesian analyses showed that the likelihood of the effects were very high; the Bayes factors indicated strong/very strong evidence for the effect of probe type on number-matching RT and substantial/strong evidence for the effect of probe type on number-matching accuracy (according to language used to report Bayes analyses that is outlined in [Jarosz & Wiley, 2014](#)).

The second finding of note concerns the prediction related to the independent vs. IE models of arithmetic fact representation. Recall that if the independent model of arithmetic fact representation were true, there would be a three-way interaction effect in the number-matching task where the probe type effect for the L x s practice group would be present (or at least, larger) for the L x s cue trials and not present (or at least, smaller) for the s x L trials, and the reverse would be true for the s x L practice group. This interaction effect did not materialize; in fact, the resulting Bayes factors from the interaction analyses ranged between .14 and .30 (specifically, .14 for both three-way interactions), which corresponds to the null model being from 3.37 to 7.14 times more likely to occur under a model that does not include the interaction effects than one that does. This is positive/substantial evidence for the null hypothesis ([Jarosz & Wiley, 2014](#)). Instead, the data fit what one would expect given the Identical Elements model, which states that the same representation is accessed/strengthened regardless of the operand order that is being practiced. This would lead to equal interference effects that do not depend on whether the order of the cue matches the order of the operands in the practiced multiplication problems.

### Experiment 3

The focus of Experiment 3 was to extend to complex multiplication another set of effects that have been documented in recent investigations of simple multiplication. According to the Interacting Neighbors (IN) Model of single-digit multiplication, how many “neighbors” an arithmetic problem has will have an impact on the speed and accuracy with which people can retrieve its answer (e.g., [Domahs et al., 2007](#); [Verguts & Fias, 2005a](#)). This is analogous to effects identified in word reading where words that have many neighbors with similar spellings but that are

pronounced differently (e.g., wear, pearl, and hear) are read more slowly and incorrectly than words that do not have neighbors that are pronounced differently (past, fast, last, etc.). It has already been established in simple multiplication that when a problem's operands ( $6 \times 3$ ) are presented, the correct answer to the problem (18) and correct answers to adjacent problems (12 and 24) are also activated (Galfano et al., 2003). This demonstrates the relatedness effect—when a problem is presented ( $6 \times 3$ ), problem nodes in a semantic field are activated to varying degrees. Those problems with an operand match that also have a second operand that differs by one or two units (e.g.,  $7 \times 3$  and  $7 \times 4$ ) are activated more strongly, while those that share an operand and have a second operand more than two units away (e.g.,  $4 \times 3$ ) or those that do not match either operand (e.g.,  $8 \times 5$ ) are activated weakly.

Other activation spreads in the semantic field as well. For example, the problem  $6 \times 3$  activates the 1-node in a decade field because there is a “1” in the decade of the correct answer. Similarly, the 8-node in the ones field is activated because there is an “8” in the ones place of the correct answer. In light of this activation of the different numerical components of answers, the IN model predicts that problems that have the most neighbors that activate the decade and unit numbers that are in the correct answer, the faster the correct answer will be retrieved. To demonstrate with the  $6 \times 3$  example, because its answer shares the decade digit with the answers 12 and 15, and because  $6 \times 3$  will activate nearby problems  $6 \times 2$ ,  $4 \times 3$ , and  $5 \times 3$ , the 1-node for the decade part of the answer will be activated highly. A problem such as  $9 \times 7$  won't have the same degree of facilitation as  $6 \times 3$  because only one problem has an answer in the same decade (64) and will only be activated weakly by  $9 \times 7$  because the problem with that answer,  $8 \times 8$ , does not share an operand with  $9 \times 7$ . When neighboring problems share decades and/or unit numbers of their answer (e.g.,  $6 \times 3 = 18$  and  $6 \times 2 = 12$ ), they are said to be “consistent;” when they do not (e.g.,  $6 \times 3 = 18$  and  $6 \times 4 = 24$ ), they are termed “inconsistent.”

Recent studies using connectionist modeling (Verguts & Fias, 2005b), production tasks in which participants verbalize answers (Domahs, Delazer, & Nuerk, 2006; Verguts & Fias, 2005a), and verification tasks in which participants decide whether a given answer is correct (Domahs et al., 2007), all have produced evidence in support of the IN model's predictions about relatedness and consistency effects. For example, Domahs et al. (2007) gave participants simple multiplication problems to verify ( $8 \times 4 = 36?$ ) while simultaneously recording ERP data. In addition to verifying correct answers ( $8 \times 4 = 32$ ), participants had to reject answers that were related and consistent ( $8 \times 4 = 36$ ), related and inconsistent ( $8 \times 4 = 28$ ), unrelated and consistent ( $8 \times 4 = 38$ ), or unrelated and inconsistent ( $8 \times 4 = 26$ ). The behavioral results indicated that across the SOAs that were used, related answers were rejected 100 ms more slowly than unrelated answers, and inconsistent answers were rejected 23 ms more slowly than consistent answers. Analysis of the error rates across conditions yielded the same pattern.

Given the review of the IN model and findings concerning relatedness and consistency effects, the purpose of Experiment 3 was to determine if the findings would extend to complex multiplication problems. To test this, participants who were not involved in the previous two experiments were asked to practice a new set of complex multiplication problems. Subsequently, a verification task was administered in which incorrect answers varied in their consistency and relatedness. The hypothesis is RTs and error rates will increase in a post-practice verification task when the answers to be verified are consistent and/or related.

## Method

### Participants

Forty undergraduates from a Northeastern college in the USA signed up for the experiment. Four of them did not return for their second appointment, leaving 36 participants (25 females and 11 males) with a mean age of 20.3 years who completed the experiment and received extra credit toward their psychology class. All participants had normal or corrected to normal vision. Three participants reported English was their second language, one participant reported a diagnosed attention disorder, and none of the participants indicated that they had a diagnosed math disability. As in the previous two experiments, data analysis including or deleting the four participants' data resulted in the same pattern of results, so their data were retained in the analyses below. The apparatus was the same as was used in Experiments 1 and 2.

**Complex multiplication practice stimuli** — A new set of six complex multiplication problems and their commuted pairs was constructed so that both relatedness and consistency could be varied systematically in the multiplication verification task described below (see [Appendix D](#) for the practice stimuli). Problems were presented in blocks of 12, and within each of the blocks, they were selected at random for presentation.

**Complex multiplication verification stimuli** — These stimuli were constructed to control many confounding variables, largely following how [Domahs et al. \(2007\)](#) created their verification stimuli. Six sets of items were composed, and each set was used twice, once for a problem and a second time for its commuted pair. Each set included a correct answer and four incorrect answers, called lures—a consistent related lure, an inconsistent related lure, a consistent unrelated lure, and an inconsistent unrelated lure. Using the problem  $4 \times 14$  as an example, a consistent-related lure was a correct answer to a near-neighbor practice problem that contained the same decade digit (e.g., 52, the correct answer to  $4 \times 13$ ); an inconsistent-related lure was a correct answer to a near-neighbor practice problem whose decade digit did not match (e.g., 60, the correct answer to  $4 \times 15$ ). A consistent-unrelated lure was a presented answer that was not an answer to a practiced problem but did share the decade digit (e.g., 58), while an inconsistent-unrelated lure was not an answer to a practiced problem and did not share the decade digit (e.g., 62). In each block, the correct answer appeared four times, and each type of lure appeared once; this balanced the number of correct and incorrect trials.

Because incorrect answers can be rejected without actually calculating a correct answer when odd-even (parity) status of problem and answer do not match and/or the split (distance from the correct answer) of the answer to be verified is large, these variables were controlled (e.g., [Lemaire & Fayol, 1995](#)). As all of the problems in the experiment contained the even single digit “4,” to preserve parity, all incorrect answers were even. On average, the splits were kept relatively equal in magnitude; for related lures the average split was 5.3 and for unrelated lures was 5.8, while for consistent lures it was 4.8 and for inconsistent lures was 6.3.

As much as possible, lures did not contain digits that matched in congruent positions with the problem (e.g.,  $8 \times 4 = 34$ ). The only exception to this was the problems  $4 \times 18$  and  $18 \times 4$ ; for these, there was a match for both the related inconsistent (68) and unrelated consistent (78) lures. Because interference due to this match is uncommon when the first operand matches a number in the answer, minimal additional interference from the problem  $18 \times 4$  to reject either of the lures should occur (e.g., [Campbell, 1997](#)). Also, given that the match occurs for both a related and an unrelated answer, and for a consistent and inconsistent answer, any potential added interference is balanced across the four answer categories.

## Procedure

The sign-up and length/scheduling of practice were the same as in Experiments 1 and 2. After practice, participants completed the complex multiplication verification task that took approximately 25 minutes. Procedures in the practice task were the same as in Experiment 1, including the instructions, completion of warm-up problems that were different from those in the data-collection portion of the task, the use of the STRATEGY screen, and the timing and sequencing of the stimuli.

For the verification task, participants were instructed to use their two index fingers to make the button-press responses to indicate an incorrect or correct answer. Participants were randomly assigned to use the left or right button to indicate a “correct” answer and the other button to indicate an “incorrect” answer. The task began with a set of instructions that a research assistant read aloud while a participant read them silently. Participants ran through a set of 36 warm-up problems to get accustomed to the verification task and the correct use of the two response buttons. The sequence and timing of the stimuli in each of the trials almost exactly mirrored those used in the long-SOA condition in Domahs et al. (2007)— see Appendix E for a visual representation of the sequence and timing of events. This was selected because Domahs et al. found significant effects for both relatedness and consistency at this SOA. First, an “X” was presented in the center of the computer screen for 300 ms to orient a participant’s attention, which was followed immediately by a white screen for 200 ms. Next, the operands of the multiplication problem, without the multiplication sign between them, were presented for 100 ms, and then a white screen was presented for 450 ms, yielding an SOA of 550 ms. Finally, an answer was presented and remained on the screen until the participant had pressed a button on the response pad to verify the answer or until 2500 ms had elapsed. Another 1000 ms blank screen appeared before the beginning of the next trial.

A total of 384 trials were presented. Of these, 192 were correct trials, and 192 were incorrect trials, the latter of which were divided evenly into 48 trials of each of the incorrect answer types. This meant that each problem and its commuted pair were presented 16 times with their correct answer and four times with each of the different types of lures. These 384 trials were separated into eight blocks; during each block, 24 correct trials (two repetitions of each commuted pair) and 24 incorrect trials (one of the commuted pairs from each lure category) were presented. After each block, participants were offered a break to ensure that visual/attentional fatigue would not compromise performance. Once the verification task was completed, the research assistant debriefed each participant before releasing him/her from the experiment.

## Results and Discussion

Response times to correct trials and error rate data were collected from participants in the verification task. In 3.0% of the trials, participants did not make a response before the 2500 ms cut-off. The no-response rate was 2.7% for correct answers as well as consistent related and inconsistent related lures; 4.0% for consistent unrelated lures; and 3.2% for inconsistent unrelated lures. As a result the hybrid RT trimming procedure, 1.9% of correct answer, and 2.0% of consistent related, 1.7% of inconsistent related, 2.0% of consistent unrelated, and 2.1% of inconsistent unrelated lure RTs were removed. Four participants were excluded from the data analysis below because their combined error and no-response rates were very high (43%, 47%, 51%, and 61%), suggesting that they were simply guessing when verifying answers.

A 2 x 2 repeated-measures ANOVA was conducted with relatedness (related vs. unrelated) and consistency (consistent vs. inconsistent) as factors. For the descriptive statistics related to this analysis, refer to Table 3. The main effect for relatedness was significant, as related lures were rejected 61 ms slower than unrelated lures,  $F(1,$

31) = 16.36,  $p < .001$ ,  $\eta^2_p = .35$ ,  $BF_{10} > 150$ . In contrast, the 27 ms longer RTs for rejecting consistent compared to inconsistent lures was not significant,  $F(1, 31) = 2.73$ ,  $p > .05$ ,  $\eta^2_p = .08$ ,  $BF_{10} = .68$ . The difference between responses to the consistent versus inconsistent lures was 12 ms longer for related versus unrelated lures, but this interaction effect was not significant,  $F(1, 31) = .24$ ,  $p > .05$ ,  $\eta^2_p < .01$ ,  $BF_{10} = .20$ .

Table 3

*Mean Multiplication Verification Response Times and Accuracies by Lure (Answer) Type*

	Correct	Related Consistent	Related Inconsistent	Unrelated Consistent	Unrelated Inconsistent
Response Time	820 (167)	958 (194)	926 (161)	891 (174)	871 (182)
Accuracy	.89 (.09)	.81 (.13)	.88 (.14)	.89 (.12)	.92 (.10)

*Note.* Response times are in milliseconds and accuracies are proportion correct. Standard deviations are in parentheses.

A second 2 x 2 ANOVA with the same factors was conducted on the accuracy data. Both of the main effects were significant. Error rates were 6% higher for related than for unrelated lures,  $F(1, 31) = 29.98$ ,  $p < .001$ ,  $\eta^2_p = .49$ ,  $BF_{10} > 150$ , and were 5% higher for consistent compared to inconsistent lures,  $F(1, 31) = 49.10$ ,  $p < .001$ ,  $\eta^2_p = .61$ ,  $BF_{10} > 150$ . Additionally, the error rate for consistent versus inconsistent lures was 3% higher for related compared to unrelated lures, which was significant,  $F(1, 31) = 4.18$ ,  $p < .05$ ,  $\eta^2_p = .12$ ,  $BF_{10} = 1.34$ . Once again, due to non-normally distributed residuals, Wilcoxon sign-rank tests were also conducted on the accuracy data yielding significant effects of relatedness,  $Z = 4.20$ ,  $p < .001$ , and consistency,  $Z = 4.47$ ,  $p < .001$ ; the interaction effect, however, did not materialize,  $Z = 1.32$ ,  $p > .05$ . The lack of an interaction effect in the non-parametric analysis isn't much at odds with the ANOVA finding, as the Bayes factor of 1.34 for the interaction effect corresponds to an anecdotal/weak effect (Jarosz & Wiley, 2014).

In summary, as predicted, after complex multiplication practice, participants were slower and more error prone when rejecting lures that were the correct answer to another practiced problem (related), and were more error prone when the lures shared the decade number with the correct answer to the problem (consistent). The sizes of the effects were fairly large, as the variance accounted for according to partial eta squares were 35% (relatedness effect for RTs), 49% (relatedness effect for accuracies), and 61% (consistency effect for accuracies). These effects, given that the Bayes factors were greater than 150, can be characterized as very strong/decisive (Jarosz & Wiley, 2014). The magnitudes of these effects were in line with those documented at the longer SOA in Domahs et al.'s (2007) study of simple multiplication, which used a similar methodology—relatedness effect 61 ms (this study) versus 74 ms (Domahs et al.) and consistency effect 27 ms (this study) versus 30 ms (Domahs et al.); unfortunately accuracy was not reported by Domahs et al. by lure type and cannot be compared. Taken together, these findings further support the Interacting Neighbors Model that posits that representations of two-digit number answers are decomposed into decades and units during arithmetic processing.

## General Discussion

Recall that the purpose of the experiments was twofold. The first was to document that the obligatory activation of arithmetic answers extends from simple to complex multiplication problems. Both Experiments 1 and 2 demonstrated that this activation does extend to participants who, through recent practice, have become skilled

at complex multiplication problem solving. As noted previously, in these two experiments, this obligatory activation was indexed using a number-matching task which does not require multiplication processing. Results showed after roughly 75 minutes of practice on a subset of problems that product probe-trial RTs/error rates were significantly longer/larger than unrelated probe-trial RTs/error rates.

The second goal of the set of experiments was to show that complex multiplication could be used to evaluate existing models of arithmetic fact representations. In Experiment 2, this was achieved by having participants practice the same subset of complex arithmetic problems with two practice groups employed—one that only saw problems with the smaller operand first ( $s \times L$ ) and the other that only saw problems with the larger operand first ( $L \times s$ ). The number-matching task was used again to determine whether the practice groups showed different levels of interference when the cues presented matched, versus did not match, their practiced order. The probe interference effects across the different cue types ( $s \times L$ ,  $L \times s$ ) did not differ as a function of practice group—in other words, no three-way interactions were found, and Bayes factors provided positive evidence for the null hypothesis. These findings support single-representation models of arithmetic that predict that practice of either problem order should strengthen the same single representation (Rickard, 2005; Rickard & Bourne, 1996). An additional interesting possibility to explore in future research is the role, if any, that perceptual factors play as Rickard and Bourne (1996) did. For example, how would changing the perceptual factors of practice (e.g., problems are presented in word format) and/or of the cue/probe format (e.g., one or both are presented in word format) affect the interference/practice effects just described? Only LeFevre et al. (1988) examined format change in a number-matching task, and it was only comparing the probe in digit versus word format, which resulted in null effects.

Turning to Experiment 3, in a multiplication verification task that followed complex multiplication practice, participants were slower and/or more error prone when they had to reject incorrect answers that shared the decade digit of the correct answer (consistency effect) or were a correct answer to a different practiced problem (relatedness effect). The replication of these effects for complex multiplication problems supports the IN Model of arithmetic that multi-digit numbers, in particular answers to multiplication problems, are processed in a decomposed (decade and units digits separately), rather than holistic fashion (Verguts & Fias, 2005a).

As mentioned previously, probe type, consistency, and relatedness effects across the three experiments were very similar to those found in simple multiplication investigations. This is particularly noteworthy given that in the present investigations the number of presentations of the complex problems was limited and occurred over a short time frame, while adults have encountered simple multiplication problems much more extensively and over a very long period of time. Recall that a medium-length SOA was used in the number-matching tasks, and the one used in the arithmetic verification task in Experiment 3 matched the longer SOA in Domahs et al. (2007). So it may be that the aforementioned effects occur only at middle/long SOAs with samples who have had few problem exposures and the practice has been massed rather than gradual. In fact, there is some evidence from the simple arithmetic literature that more-skilled participants show significant number-matching interference at shorter SOAs but not longer ones and that the reverse is true for less-skilled participants (LeFevre, Kulak, & Bisanz, 1991); also, the size of the effect may be dependent on participant skill and SOA as well, as the participants in Domahs et al. (2007) showed larger relatedness effects and smaller consistency effects at the longer SOA, the latter effect probably indicating that consistency effects follow a slower time course. It would be interesting to conduct a longitudinal practice study to test how the above effects interact with SOA for complex multiplication problems as practice parameters (number of problem exposures and overall practice time-frame) are varied.

Give the findings of the three experiments, follow-up experiments using complex multiplication practice and number-matching could be used to further our understanding of how arithmetic problems and their answers are represented in long-term memory. For example, recently there has been a proliferation of preferred-representation models of arithmetic (a different type of single-representation model). These suggest that addition and multiplication problems may have a more privileged addend/operand order representation in long-term memory that corresponds to speed and error rate differences in solving problems presented in the preferred versus less-preferred operand order (e.g., [Didino, Lombardi, & Vespignani, 2014](#); [Zhou, Zhao, Chaunsheng, & Zhou, 2012](#)).

For example, a recent set of experiments showed that Italian adults solved  $L \times s$  problems faster than  $s \times L$  problems but only when at least one of the operands was smaller than five; when both were larger than five,  $s \times L$  problems were solved faster ([Didino et al., 2014](#)). This was explained to have occurred as a combined result of strategy use and learning order. A very common strategy used when first solving multiplication problems is repeated addition ( $3 \times 7$  is  $7 + 7 + 7 = 21$ ). When at least one number is smaller than five, this is a fairly efficient strategy to use because at most, one needs only to add four numbers. When the problem is presented in the  $s \times L$  format, it is reversed to the  $L \times s$  format so that a person may execute the repeated addition strategy efficiently. Because the repeated addition strategy overall is much less efficient for solving problems that have both operands greater than five (minimally, one repeatedly has to add a large number six times), this strategy is not used, and what dominates which operand order is stored most strongly in long-term memory is one's learning history. Problems are presented in the  $s \times L$  format first in Italy, so that is the preferred representation for problems with both operands larger than five.

There are two issues that cloud the results and related causal explanations of [Didino et al. \(2014\)](#). First, they administered a strategy assessment task in their experiments and found that the use of non-retrieval strategies was over 45%. The strategy data also indicated that retrieval was used more frequently for small (both operands less than five) and medium problems (one operand less than five) when presented in the  $s \times L$  order versus the  $L \times s$  order, and the opposite was true for the large problems (both operands larger than five). The assumption that the solution procedures used for each problem in the strategy assessment (no time pressure) would be used for the same problems in the chronometric condition (speeded responding) is debatable ([Campbell & Austen, 2002](#)). Even if the assumption is valid, however, the RT differences could be due to differences in non-retrieval strategy use for  $s \times L$  versus  $L \times s$  problems rather than differences in speed of accessing the answers from long-term memory; [Didino et al.](#) did note this possibility in the discussion of their results.

A second issue is the repeated addition strategy and learning order explanations offered by [Didino et al. \(2014\)](#) to explain the interaction between operand order and problem size; they are intriguing, but they would be more compelling if they were evaluated experimentally. Both of the aforementioned issues could be addressed using a design similar to what was used in Experiment 2 of the present investigation. In a future experiment, both strategy use and order of learning problems could be varied. Groups of participants could be assigned to practice a subset of complex multiplication problems using different strategies, and problems also could be introduced in different orders (e.g., one group practices  $s \times L$  problems first, the other practices  $L \times s$  problems first); this would enable experimenters to determine the effects that both strategy use and learning order has on the subsequent representation of problems and answers in long-term memory. After practice, both arithmetic production and verification tasks could be administered along with a number-matching task. Using the number-matching task would provide a pure measure of the automatic retrieval activation of the different operand orders—no assumptions

or measurement of strategy use would need to be made. Results using the number-matching task also could be compared to those found in the production and verification tasks to see if they converge.

In conclusion, the set of experiments conducted here have shown that the use of a complex multiplication practice design along with number-matching tasks have extended several well documented effects from the simple multiplication literature. Designing future experiments that employ complex arithmetic practice as a component should assist researchers in evaluating models of arithmetic processing. Practice experiments using pseudo number-arithmetic tasks have been conducted previously to evaluate arithmetic models (alphaplication such as “I, E = p,” e.g., [Graham & Campbell, 1992](#); diamond arithmetic, e.g., [Whalen, 1997](#)), and while these tasks share some important structural and conceptual features with number-arithmetic, they diverge in important ways as well. Given that most individuals have had little experience mentally solving complex problems, use of designs involving complex multiplication practice enables researchers to control variables that can’t be controlled in simple arithmetic studies such as strategy use, problem presentation order, and problem type (e.g., low vs. high neighborhood density) while avoiding the drawbacks of using pseudo number-arithmetic tasks.

## Funding

The author has no funding to report.

## Competing Interests

The author has declared that no competing interests exist.

## Acknowledgments

I would like to thank the following undergraduate research assistants who were instrumental in collecting data for the three experiments: Alexia Levin, Andrea DeCusati, Anthony Corvino, Amy Benvenuto, Tiffany “Tippy” Shaffer, and Rob Gambardella. Further, I would like to thank Dr. John Towse and two anonymous reviewers for their incredibly helpful comments that significantly improved the quality of the manuscript.

## References

- Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*, 75-106. doi:[10.1016/0010-0277\(92\)90051-I](https://doi.org/10.1016/0010-0277(92)90051-I)
- Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition*, *53*, 1-44. doi:[10.1016/0010-0277\(94\)90075-2](https://doi.org/10.1016/0010-0277(94)90075-2)
- Campbell, J. I. D. (1995). Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Mathematical Cognition*, *1*, 121-164. doi:[10.1080/135467999387270](https://doi.org/10.1080/135467999387270)
- Campbell, J. I. D. (1997). Reading-based interference in cognitive arithmetic. *Canadian Journal of Experimental Psychology*, *51*, 74-81. doi:[10.1037/1196-1961.51.1.74](https://doi.org/10.1037/1196-1961.51.1.74)
- Campbell, J. I. D., & Alberts, N. M. (2009). Operation-specific effects of numerical surface form on arithmetic strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 999-1011. doi:[10.1037/a0015829](https://doi.org/10.1037/a0015829)
- Campbell, J. I. D., & Austen, S. (2002). Effects of response time deadlines on adults’ strategy choices for simple addition. *Memory & Cognition*, *30*, 988-994. doi:[10.3758/BF03195782](https://doi.org/10.3758/BF03195782)

- Campbell, J. I. D., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, *44*, 1255-1265. doi:10.3758/s13428-012-0186-0
- De Brauwer, J. L. (2007). *The representations of arithmetic fact knowledge in memory and their development* (Doctoral dissertation, Ghent University, Ghent, Belgium). Retrieved from <http://hdl.handle.net/1854/LU-468885>
- Dehaene, S., & Akhaverin, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 314-326. doi:10.1037/0278-7393.21.2.314
- Didino, D., Lombardi, L., & Vespignani, F. (2014). Operand-order effect in multiplication and addition. *Experimental Psychology*, *61*, 470-479. doi:10.1027/1618-3169/a000264
- Domahs, F., Delazer, M., & Nuerk, H. C. (2006). What makes multiplication facts difficult: Problem size or neighborhood consistency? *Experimental Psychology*, *53*, 275-282. doi:10.1027/1618-3169.53.4.275
- Domahs, F., Domahs, U., Schlesewsky, M., Ratinckx, E., Verguts, T., Wilmes, K., & Nuerk, H.-C. (2007). Neighborhood consistency in mental arithmetic: Behavioral and ERP evidence. *Behavioral and Brain Functions*, *3*, 66-78. doi:10.1186/1744-9081-3-66
- Fanget, M., Thevenot, C., Castel, C., & Fayol, M. (2015). Retrieval from memory or procedural strategies for addition problems: The use of the operand-recognition paradigm in 10-year-old children. *Swiss Journal of Psychology*, *70*, 35-39. doi:10.1024/1421-0185/a000036
- Galfano, G., Mazza, V., Angrilli, A., & Umiltà, C. (2004). Electrophysiological correlates of stimulus-driven multiplication facts retrieval. *Neuropsychologia*, *42*, 1370-1382. doi:10.1016/j.neuropsychologia.2004.02.010
- Galfano, G., Penolazzi, B., Vervaeck, I., Angrilli, A., & Umiltà, C. (2009). Event-related brain potentials uncover activation dynamics in the lexicon of multiplication facts. *Cortex*, *45*, 1167-1177. doi:10.1016/j.cortex.2008.09.003
- Galfano, G., Rusconi, E., & Umiltà, C. (2003). Automatic activation of multiplication facts: Evidence from the nodes adjacent to the product. *The Quarterly Journal of Experimental Psychology*, *56A*, 31-61. doi:10.1080/02724980244000332
- Geary, D. C. (1996). The problem-size effect in mental addition: Developmental and cross-national trends. *Mathematical Cognition*, *2*, 63-94. doi:10.1080/135467996387543
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge of children with mathematical disability. *Journal of Experimental Child Psychology*, *88*, 121-151. doi:10.1016/j.jecp.2004.03.002
- Grabner, R. H., Ischebeck, A., Reishofer, G., Koschutnig, K., Delazer, M., Ebner, F., & Neuper, C. (2009). Fact learning in complex arithmetic and figural-spatial tasks: The role of the angular gyrus and its relation to mathematical competence. *Human Brain Mapping*, *30*, 2936-2952. doi:10.1002/hbm.20720
- Grabner, R. H., Rüttsche, B., Ruff, C. C., & Hauser, T. U. (2015). Transcranial direct current stimulation of the posterior parietal cortex modulates arithmetic learning. *The European Journal of Neuroscience*, *42*, 1667-1674. doi:10.1111/ejn.12947
- Graham, D. J., & Campbell, J. I. D. (1992). Network interference and number-fact retrieval: Evidence from children's alphaplication. *Canadian Journal of Psychology*, *46*, 65-91. doi:10.1037/h0084310

- Hittmair-Delazer, M., Semenza, C., & Denes, G. (1994). Concepts and facts in calculation. *Brain*, *117*, 715-728. doi:10.1093/brain/117.4.715
- Hoard, M. K., Geary, D. C., Byrd-Craven, J., & Nugent, L. (2008). Mathematical cognition in intellectually precocious first graders. *Developmental Neuropsychology*, *33*, 251-276. doi:10.1080/87565640801982338
- Imbo, I., & LeFevre, J.-A. (2010). The role of phonological and visual working memory in complex arithmetic for Chinese- and Canadian-educated adults. *Memory & Cognition*, *38*, 176-185. doi:10.3758/MC.38.2.176
- Imbo, I., & Vandierendonck, A. (2008). Effects of problem size, operation, and working-memory span on simple-arithmetic strategies: Differences between children and adults? *Psychological Research*, *72*, 331-346. doi:10.1007/s00426-007-0112-8
- Ischebeck, A., Zamarian, L., Egger, K., Schocke, M., & Delazer, M. (2007). Imaging early practice effects in arithmetic. *NeuroImage*, *36*, 993-1003. doi:10.1016/j.neuroimage.2007.03.051
- Ischebeck, A., Zamarian, L., Schocke, M., & Delazer, M. (2009). Flexible transfer of knowledge in mental arithmetic—An fMRI study. *NeuroImage*, *44*, 1103-1112. doi:10.1016/j.neuroimage.2008.10.025
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *Journal of Problem Solving*, *7*, 2-9. doi:10.7771/1932-6246.1167
- Kirk, E. P., & Ashcraft, M. H. (2001). Telling stories: The perils and promise of using verbal reports to study math strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 157-175. doi:10.1037/0278-7393.27.1.157
- LeFevre, J.-A., 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. doi:10.1037/0096-3445.125.3.284
- LeFevre, J.-A., Bisanz, J., & Mrkonjic, L. (1988). Cognitive arithmetic: Evidence for obligatory activation of arithmetic facts. *Memory & Cognition*, *16*, 45-53. doi:10.3758/BF03197744
- LeFevre, J.-A., Kulak, A. G., & Bisanz, J. (1991). Individual differences and developmental change in the associative relations among numbers. *Journal of Experimental Child Psychology*, *52*, 256-274. doi:10.1016/0022-0965(91)90062-W
- Lemaire, P., & Fayol, M. (1995). When plausibility judgments supersede fact retrieval: The example of the odd-even effect on product verification. *Memory & Cognition*, *23*, 34-48. doi:10.3758/BF03210555
- Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, *43*, 679-690. doi:10.3758/s13428-010-0049-5
- McCloskey, M., Aliminos, D., & Sokol, S. M. (1991). Facts, rules, and procedures in normal calculation: Evidence from multiple single-case studies of impaired arithmetic fact retrieval. *Brain and Cognition*, *17*, 154-203. doi:10.1016/0278-2626(91)90074-I
- Pesenti, M., Seron, X., Samson, D., & Duroux, B. (1999). Basic and exceptional calculation abilities in a calculating prodigy: A case study. *Mathematical Cognition*, *5*, 97-148. doi:10.1080/135467999387270
- R Development Core Team. (2012). *R: A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria*. Retrieved from <http://www.R-project.org>

- Rickard, T. C. (2005). A revised identical elements model of arithmetic fact representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 250-257. doi:10.1037/0278-7393.31.2.250
- Rickard, T. C., & Bourne, L. E., Jr. (1996). Some tests of an identical elements model of basic arithmetic skills. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1281-1295. doi:10.1037/0278-7393.22.5.1281
- Rickard, T. C., Healy, A. F., & Bourne, L. E., Jr. (1994). On the cognitive structure of basic arithmetic skills: Operation, order, and symbol transfer effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1139-1153. doi:10.1037/0278-7393.20.5.1139
- Rusconi, E., Galfano, G., Rebonato, E., & Umiltà, C. (2006). Bidirectional links in the network of multiplication facts. *Psychological Research*, 70, 32-42. doi:10.1007/s00426-004-0187-4
- Rusconi, E., Galfano, G., Speriani, V., & Umiltà, C. (2004). Capacity and contextual constraints on product activation: Evidence from task-irrelevant fact retrieval. *The Quarterly Journal of Experimental Psychology*, 57A, 1485-1511. doi:10.1080/02724980343000873
- Thevenot, C., Castel, C., Fanget, M., & Fayol, M. (2010). Mental subtraction in high and lower-skilled arithmetic problem solvers: Verbal report vs. operand-recognition paradigms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1242-1255. doi:10.1037/a0020447
- Thevenot, C., Fanget, M., & Fayol, M. (2007). Retrieval or non-retrieval strategies in mental arithmetic? An operand-recognition paradigm. *Memory & Cognition*, 35, 1344-1352. doi:10.3758/BF03193606
- Thibodeau, M. H., LeFevre, J.-A., & Bisanz, J. (1996). The extension of the interference effect to multiplication. *Canadian Journal of Experimental Psychology*, 50, 393-396. doi:10.1037/1196-1961.50.4.393
- Tronsky, L. N. (2005). Strategy use, the development of automaticity, and working memory involvement in complex multiplication. *Memory & Cognition*, 33, 927-940. doi:10.3758/BF03193086
- van Opstal, F., & Verguts, T. (2011). The origins of the numerical distance effect: The same-different task. *Journal of Cognitive Psychology*, 23, 112-120. doi:10.1080/20445911.2011.466796
- Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology*, 47A(3), 631-650. doi:10.1080/14640749408401131
- Verguts, T., & Fias, W. (2005a). Neighborhood effects in mental arithmetic. *Psychological Science*, 47, 132-140. [http://www.pabst-publishers.de/psychology-science/1-2005/ps\\_1\\_2005\\_132-140.pdf](http://www.pabst-publishers.de/psychology-science/1-2005/ps_1_2005_132-140.pdf)
- Verguts, T., & Fias, W. (2005b). Interacting neighbors: A connectionist model of retrieval in single-digit multiplication. *Memory & Cognition*, 33, 1-16. doi:10.3758/BF03195293
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of  $p$  values. *Psychonomic Bulletin & Review*, 14, 779-804. doi:10.3758/BF03194105
- Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science*, 25, 169-176. doi:10.1177/0963721416643289

Whalen, J. (1997). The influence of numerical magnitude representations on arithmetic fact retrieval: Theory, data, and simulation. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the nineteenth annual conference of the cognitive science society* (pp. 814-819). Mahwah, NJ, USA: Erlbaum.

Zhou, F., Zhao, Q., Chaunsheng, C., & Zhou, X. (2012). Mental representations of arithmetic facts: Evidence from eye movement recordings supports the preferred operand-order-specific representation hypothesis. *The Quarterly Journal of Experimental Psychology*, 65, 661-674. doi:10.1080/17470218.2011.616213

## Appendices

### Appendix A

Table A.1

*Number Matching Task Stimuli Used in Experiment 1*

<b>Non-matching</b>										
<b>Product</b>										
Cue	16 3	18 3	17 4	19 4	14 6	3 16	3 18	4 17	4 19	6 14
Probe	48	54	68	76	84	48	54	68	76	84
<b>Unrelated</b>										
Cue	16 3	18 3	17 4	19 4	14 6	3 16	3 18	4 17	4 19	6 14
Probe	52	46	92	58	72	52	46	92	58	72
<b>Fillers</b>										
Cue	67 8	34 9	23 7	54 8	72 3	8 67	9 34	7 23	8 54	3 72
Probe	58	68	46	36	82	58	68	46	36	82
<b>Matching</b>										
<b>Probe-balancing</b>										
Cue	48 3	54 3	68 4	76 4	84 6	3 48	3 54	4 68	4 76	6 84
Probe	48	54	68	76	84	48	54	68	76	84
<b>Cue-balancing</b>										
Cue	16 3	18 3	17 4	19 4	14 6	3 16	3 18	4 17	4 19	6 14
Probe	3	3	4	4	6	3	3	4	4	6
<b>Fillers</b>										
Cue	86 7	98 6	62 3	46 9	38 6	7 86	6 98	3 62	9 46	6 38
Probe	86	98	62	46	38	86	98	62	46	38

## Appendix B

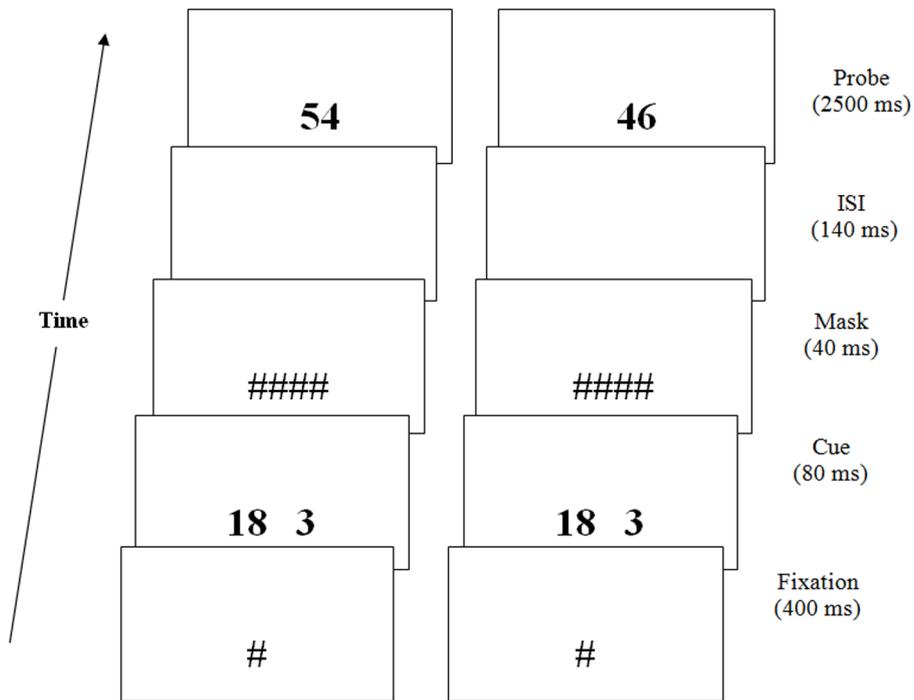


Figure B.1. Sequence of events in the number-matching task used in both Experiments 1 and 2.

## Appendix C

Table C.1

Number Matching Task Stimuli Used in Experiment 2

<b>Non-matching</b>										
<b>Product</b>										
Cue	16 3	18 3	17 4	19 4	15 6	3 16	3 18	4 17	4 19	6 15
Probe	48	54	68	76	90	48	54	68	76	90
<b>Unrelated</b>										
Cue	16 3	18 3	17 4	19 4	15 6	3 16	3 18	4 17	4 19	6 15
Probe	52	46	92	58	72	52	46	92	58	72
<b>Fillers</b>										
Cue	67 8	34 9	23 7	54 8	72 3	8 67	9 34	7 23	8 54	3 72
Probe	58	68	46	36	82	58	68	46	36	82
<b>Matching</b>										
<b>Probe-balancing</b>										
Cue	48 3	54 3	68 4	76 4	90 6	3 48	3 54	4 68	4 76	6 90
Probe	48	54	68	76	90	48	54	68	76	90
<b>Cue-balancing</b>										
Cue	16 3	18 3	17 4	19 4	15 6	3 16	3 18	4 17	4 19	6 15
Probe	3	3	4	4	6	3	3	4	4	6
<b>Fillers</b>										
Cue	86 7	98 6	62 3	46 9	38 6	7 86	6 98	3 62	9 46	6 38
Probe	86	98	62	46	38	86	98	62	46	38

## Appendix D

Table D.1

Verification Task Stimuli Used in Experiment 3

<b>Practiced Problems</b>	<b>Proposed Answer</b>				
	<b>Correct Product</b>	<b>Related-consistent</b>	<b>Related-inconsistent</b>	<b>Unrelated-consistent</b>	<b>Unrelated-inconsistent</b>
13 x 4 and 4 x 13	52	56	60	58	48
14 x 4 and 4 x 14	56	52	60	50	62
15 x 4 and 4 x 15	60	68	56	66	48
17 x 4 and 4 x 17	68	60	72	66	80
18 x 4 and 4 x 18	72	76	68	78	66
19 x 4 and 4 x 19	76	72	68	78	80

## Appendix E

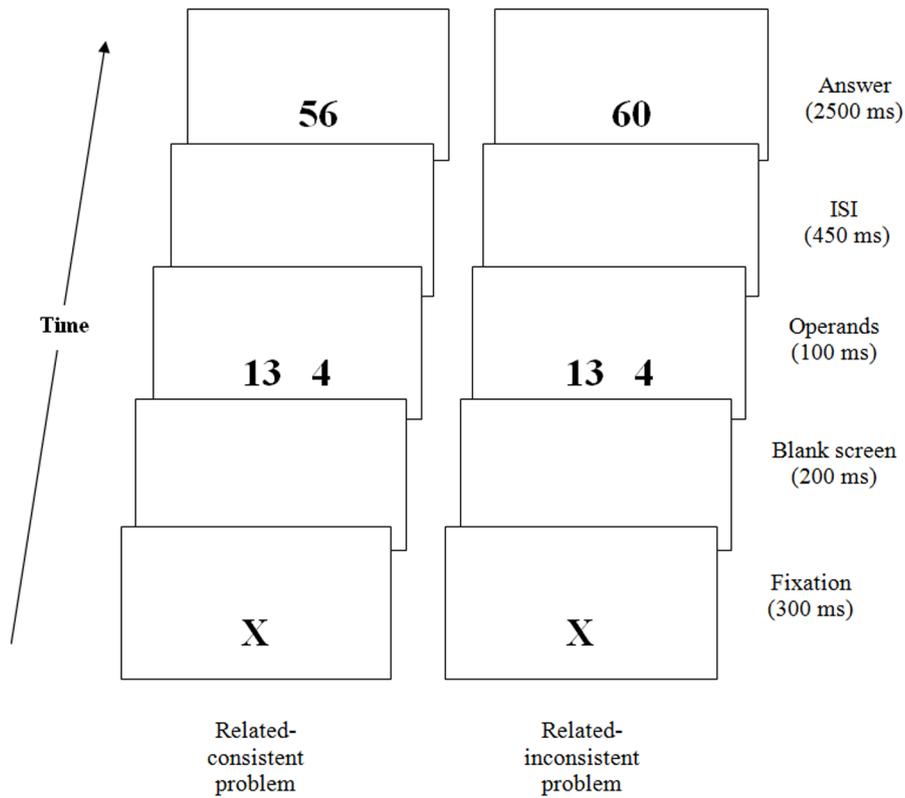


Figure E.1. Sequence of events in the verification task used in Experiment 3.