

## Empirical Research

**Reducing Interference Improves the Memorization of Multiplication Facts in Case of Hypersensitivity to Interference**Dror Dotan<sup>\*abc</sup>, Naama Friedmann<sup>b</sup>**[a]** Mathematical Thinking Lab, School of Education and the Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel.**[b]** Language and Brain Lab, School of Education and the Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel. **[c]** Cognitive Neuroimaging Unit, CEA DRF/I2BM, Neurospin, INSERM, Université Paris-Sud, Université Paris-Saclay, Gif/Yvette, France.**Abstract**

Hypersensitivity to interference (HYSTI) is a situation in which a person has a severe difficulty in memorizing verbal items that are similar to each other. This may result in induced dyscalculia: HYSTI was shown to correlate with a difficulty in learning the multiplication table, presumably because the multiplication table, which is memorized verbally, has much similarity between the items ("six times seven equals forty two", "six times eight equals forty eight", etc.). Here, we show causal evidence that HYSTI disrupts the memorization of multiplication facts. We report DL, a woman with HYSTI who had extremely poor knowledge of the multiplication table. To examine whether her multiplication difficulty resulted from HYSTI, we tested whether she could learn multiplication facts when interference was reduced. In a series of merely 12 short sessions over a period of 4 weeks, DL rehearsed 16 multiplication facts – four facts per week. When the 4 facts in a given week were similar to each other, DL's learning was poor. Conversely, when the 4 facts in a given week were dissimilar from each other, DL learned them quickly and easily. The effect of similarity was observed during the training period and persisted at least two months after the end of training. These results provide the first causal evidence that HYSTI impairs the learning or retrieval of arithmetic facts. From a pedagogical perspective, our findings may call for re-considering how multiplication facts should be taught in elementary school.

**Keywords:** hypersensitivity to interference, dyscalculia, memory, rehabilitation

Journal of Numerical Cognition, 2019, Vol. 5(3), 400–430, <https://doi.org/10.5964/jnc.v5i3.203>

Received: 2018-10-23. Accepted: 2019-02-15. Published (VoR): 2019-12-20.

\*Corresponding author at: Mathematical Thinking Lab, School of Education and the Sagol School of Neuroscience, Tel Aviv University, Tel Aviv 69978, Israel. Phone: +972-3-6408629. E-mail: dotandro@mail.tau.ac.il



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

Why is it so hard for some people to learn the multiplication table – the single-digit multiplications up to  $10 \times 10$ ? One account of such difficulties emphasizes the role of interference. The multiplication facts, which involve only 10 digits, are highly similar to each other, and this similarity may make them interfere with each other. This idea is central in the *network interference model* (Campbell, 1995; Campbell & Graham, 1985). According to this model, arithmetic facts are stored in memory as an associative network: each arithmetic fact, say  $6 \times 7 = 42$ , is associated not only with the correct solution, but also with the solutions of other facts (e.g., with 48, the product of  $6 \times 8$ ). Such incorrect-solution association may occur, for example, if the two facts share an operand (e.g.,  $6 \times 7 = 42$  and  $6 \times 8 = 48$ ). Incorrect-solution associations disrupt learning and can cause retrieval errors, because they create interference among facts: when presented with a particular fact, the person may follow the association network to the solution of another fact.

Several studies indicate that the representation underlying multiplication facts is verbal (Dehaene, 1992; Dehaene & Cohen, 1995; Dehaene, Piazza, Pinel, & Cohen, 2003). This conclusion is supported by neuropsychological studies showing that multiplication deficits are associated with verbal impairments (as opposed to subtraction deficits, which are associated with impaired quantity processing; Cohen & Dehaene, 2000; Cohen, Dehaene, Chochon, Lehéricy, & Naccache, 2000; Dagenbach & McCloskey, 1992; Dehaene & Cohen, 1997; Delazer & Benke, 1997; Lampl, Eshel, Gilad, & Sarova-Pinhas, 1994; Lochy, Domahs, Bartha, & Delazer, 2004; Pesenti, Seron, & van der Linden, 1994; van Harskamp & Cipolotti, 2001). It is also supported by brain imaging studies showing that multiplication tasks activate brain regions that are also activated by tasks of language, verbal short-term memory, and phonological processing (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Simon, Mangin, Cohen, Le Bihan, & Dehaene, 2002). There is also evidence for nonverbal multiplication (McCrink & Spelke, 2010), but the common assumption is still that by and large, multiplication is driven mostly by verbal representations. Thus, the interference effects in multiplication may be related to general mechanisms of verbal memory. Indeed, verbal memory is susceptible to interference: similar verbal items interfere with one another more than dissimilar items in a variety of contexts – in verbal memory tasks (Baddeley, 1966, 2003; Hall, 1971; Nelson, Brooks, & Borden, 1974; Oberauer & Kliegl, 2006; Oberauer & Lange, 2008; Vallar, 2006), when learning words in a new language (Pajak, Creel, & Levy, 2016), and even when memorizing non-numeric arithmetic-like facts (Graham & Campbell, 1992).

Everybody experiences some difficulties arising from verbal interference, but for some people the difficulty is more severe than for others. De Visscher and Noël (2013, 2014a, 2014b) proposed that some people have *hypersensitivity to interference* – an extreme sensitivity to the interference arising from similarity between verbal items – and that these people may experience difficulty in handling the multiplication facts. In support of this hypothesis they reported DB, a woman with poor memory of the multiplication table, and showed that she also had hypersensitivity to interference: she performed poorly in tasks that were sensitive to interference, even when the stimuli were non-number words. In contrast, she performed well in tasks that assessed several other potential sources of difficulty in calculation, including verbal working memory capacity. De Visscher and Noël proposed that DB's difficulty in memorizing the multiplication table was a reflection of a more general difficulty in verbal memory – her hypersensitivity to interference.

De Visscher and Noël's series of studies showed a convincing correlation between hypersensitivity to interference and difficulty in arithmetic facts. Here, we wish to strengthen their point by providing *causal* evidence that hypersensitivity to interference disrupts the memorization of multiplication facts. We examined DL – a woman who, similar to the woman described De Visscher and Noël, had poor memory of multiplication facts as well as hypersensitivity to interference. To show that DL's hypersensitivity to interference not only correlates with her difficulty in multiplication facts but is also the *reason* for this difficulty, we designed an experiment in which we taught her multiplication facts in a low-interference condition. We predicted that in this condition, interference will play only a minor role, and consequently DL will succeed learning the multiplication facts. In contrast, in a high-interference condition she should still have difficulty. As we shall see, this was indeed the case.

Our method to reduce interference is rooted in Campbell and Graham's (1985) network interference model. The model assumes that multiplication facts are represented as a network of associations and that different facts may use intersecting network sections, so when the activation region of one problem intersects that of another problem, retrieving the answer to the first problem increases the probability that this answer would be incorrectly retrieved for the second problem. This idea is supported by the *error priming* phenomenon: Campbell (1987)

asked participants to solve a given set of multiplication problems; when the experimental block included, on top of these problems, some additional, associated problems, the participants responded more slowly and made more errors than when the block did not include such associated problems.

Similar to Campbell (1987), here we manipulated the degree of interference by controlling the similarity between the multiplication facts presented during a particular experimental session. Based on previous findings (Campbell, 1987; De Visscher & Noël, 2013, 2014a, 2014b; Girelli, Delazer, Semenza, & Denes, 1996), we reasoned that even if hypersensitivity to interference impairs DL's ability to memorize *similar* multiplication facts, she would still be able to memorize a set of *dissimilar* multiplication facts. Importantly, although the multiplication table as a whole has much similarity between the facts, some facts are dissimilar from each other (e.g.,  $9 \times 9 = 63$  and  $7 \times 4 = 28$ ), and we hoped that DL would be able to learn such subsets of facts. Moreover, we assumed that fact A would interfere with a similar fact B if the two facts are presented within a short time from each other, but not if they are presented with sufficient temporal delay between them (this was essentially the finding of Campbell, 1987). This second assumption opens the door to teaching DL not only *dissimilar* multiplication facts but also *similar* facts: all we need to do is to present similar facts in different learning sessions, which are separated from each other by a sufficient delay.

These two foundations led to the following simple training method. We identified the multiplication facts that DL did not know, and grouped them into small sets of facts. Crucially, we constructed the different sets of facts such that they had different levels of between-item similarity within the set, i.e., different levels of induced interference. Each set was taught during one week, and importantly, during that week DL refrained from rehearsing facts from any other set. We predicted that DL would have difficulty learning the multiplication facts in high-similarity sets, but would succeed learning low-similarity sets.

## General Method

### Participant

DL was a 40-year-old woman who arrived in our lab as a potential participant in an experiment about number reading difficulties. An initial examination showed that she did not have number reading difficulties, but her knowledge of the multiplication table was severely impaired. Put in her own words, she was "clueless in multiplication". She reported that her difficulties began in elementary school, and persisted in spite of several years of hard work and private tutoring in math.

### Procedure

All tasks were administered in Hebrew. The assessment tasks were done in a quiet room in our lab or in DL's home. The training was done over the telephone, while DL was in a quiet room in her home. There was no limit on the response duration in any task (but extremely slow responses were classified as errors, as detailed below).

## Statistical Analyses

### Comparing DL With Control Participants

Statistical comparisons of DL's performance to control groups were done using Crawford and Garthwaite's (2002) one-tailed  $t$ -test, with effect size defined as  $z_{cc}$  (CC stands for case-controls) – DL's score normalized according to the control group distribution (Crawford, Garthwaite, & Porter, 2010). Control participants with outlier error rates (higher than the 75<sup>th</sup> percentile by more than 150% the inter-quartile range) were excluded. When comparing DL's performance versus the control group we typically report one-tail  $p$  values, reflecting the assumption that DL may have some cognitive deficit underlying her difficulty in multiplication, and such deficit should be manifested in poorer performance than the control group.

### Comparing Single-Subject Data Between Two Conditions

One-tailed comparisons of DL's performance between two conditions were done using a bootstrapping method. The effect we examine was indexed as the difference between the two conditions in the mean score across items:

$$\text{ScoreDiff} = (\text{mean scores of condition 1}) - (\text{mean scores of condition 2})$$

One-tailed  $p$  values were obtained by comparing the observed *ScoreDiff* value versus the random distribution of *ScoreDiff* under the null hypothesis that the two conditions do not differ. This random distribution was generated by arbitrarily classifying all scores into two conditions 10,000 times, and computing the value of *ScoreDiff* for each such random classification. For a paired analysis, in which each tested item has one score in condition 1 and one score in condition 2, each random classification was created by arbitrarily labeling each item's two scores as "condition 1" and "condition 2". For an unpaired analysis, each random classification was created by reshuffling the scores of the two conditions into two sets arbitrarily labeled "condition 1" and "condition 2", while maintaining the original number of items in each condition. When the total number of items was too small such that there were fewer than 10,000 different possible random classifications, we computed the random distribution based on all possible classifications. As effect size we report  $z$  – the observed *ScoreDiff* value, standardized to the mean and standard deviation of the random distribution.

## Initial Cognitive Assessment

### Basic Arithmetic

In a screening test, DL was presented with 12 multiplication facts: 4 rule-based facts ( $N \times 0$  and  $N \times 1$ ) and 8 non-rule facts (both operands between 2 and 9). She had difficulty in both types of facts (Table 1). In contrast, she performed well in several other oral tests of arithmetic: single-digit additions, subtractions with single-digit second operand and single-digit result, and two-digit calculations<sup>1</sup>. Thus, her difficulty in multiplication facts did not disrupt addition/subtraction facts or the execution of calculation algorithms. At this time, we taught her the rules  $N \times 0 = 0$  and  $N \times 1 = N$ . Throughout the study, she hardly ever again erred in these rules.

Table 1

*DL's Performance in the Screening Calculation Tasks*

Task	No. of items	Errors	Hesitations
1-digit multiplication (rule-based)	4	2	1
1-digit multiplication (non-rule)	8	2	2
1-digit addition	15	0	0
1-digit subtraction	8	0	0
2-digit addition	3	0	0
2-digit subtraction	3	0	0
2-digit multiplication	3	0	0

We then tested DL's knowledge of all 55 multiplication facts (the larger operand always appeared first) – 19 rule-based facts and 36 non-rule facts<sup>ii</sup>. She was flawless in the rule-based facts, but she had 14/36 errors (39%) in the non-rule facts: she made 5 operator errors (adding instead of multiplying), 2 within-table errors (saying the result of another multiplication fact), 1 out-of-table error (saying a number that is not the product of any two digits), and 7 “don't know” responses. Her error rate of 39% was significantly worse than the performance of 12 age-matched control participants ( $M = 2.5$  errors,  $SD = 2.02$ , Crawford and Garthwaite's (2002)  $t(11) = 5.47$ ,  $p < .001$ ,  $z_{cc} = 5.69$ ; control group mean age = 40;7,  $SD = 5;3$ ), and worse than the worst-performing control participant, who had only 4/36 errors (using the paired bootstrap method described in General Method, one-tailed  $p = .006$ ,  $z = 2.12$ ). Thus, DL clearly had impaired knowledge of the multiplication table. We next examined potential origins for this impairment.

## Cognitive Assessment

We assessed DL's language abilities, number processing, and verbal memory, in order to examine whether her difficulty in multiplication could be explained by a deficit in any of these mechanisms. As we shall see, DL performed well in all tasks except one.

### Reading, Lexical Retrieval, and Symbolic Number Processing

DL performed well in reading words, nonwords, and word pairs (TILTAN battery, Friedmann & Gvion, 2003a), as well as in picture naming (SHEMESH test, Biran & Friedmann, 2004) (not significantly different from controls, Crawford and Garthwaite's one-tailed  $t$ -test; Table 2). She also performed well in processing symbolic numbers (digit strings and number words): reading aloud from paper a list of multi-digit numbers with 3-6 digits, repeating the same numbers, and writing numbers (3-5 digits) to dictation as digit strings. Thus, DL had good reading, lexical retrieval, and processing of symbolic numbers. Her difficulty in memorizing the multiplication facts did not originate in any of these processes.

Table 2

*DL's Performance (Either Span or Percentage of Correct Responses) in Tasks That Assess Possible Origins of her Difficulty in Multiplication Facts. She showed good reading, lexical retrieval, symbolic number processing, and memory spans. She had a difficulty only in the 2-back task*

Task	No. of items	DL	Control group			DL vs. controls (one-tailed $p$ ) <sup>a</sup>			
			<i>M</i>	<i>SD</i>	<i>n</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>z<sub>cc</sub></i>
<b>Reading</b>									
Single words	136	100%	98.3%	1.5%	372	–			
Nonwords	40	100%	95.9%	4.2%	372	–			
Word pairs	30×2	96.7%	97.5%	2.4%	372	0.33	371	.37	0.50
<b>Lexical retrieval</b>									
Picture naming	100	97.0%	97.7%	1.7%	87	0.41	86	.34	0.41
<b>Symbolic number processing</b>									
Reading aloud	120	96.7%	97.2%	1.3%	21	0.38	20	.36	0.38
Repetition	120	91.0%	95.4%	3.6%	20	1.25	19	.11	1.22
Dictation	68	97.1%	97.5%	2.3%	20	0.21	19	.42	0.17
<b>Verbal-phonological short-term memory: phonological input buffer</b>									
Word matching span <sup>b</sup>		5	6.33	0.98	12	1.3	11	.11	1.36
Digit matching span <sup>b</sup>		7	7	0	10	–			
<b>Verbal-phonological short-term memory: phonological output buffer</b>									
Digit span <sup>b</sup>		7	7.05	1.28	29	0.04	28	.49	0.04
Word span <sup>b</sup>		6	5.57	0.75	35	–			
Nonword span <sup>b</sup>		3	3.46	0.54	37	0.84	36	.20	0.85
Nonword reading	40	100%	95.9%	4.2%	372	–			
Nonword repetition	48	97.9%	95.4%	3.5%	20	–			
<b>Working memory</b>									
Digit span backward <sup>c</sup>									
			Max. level = 8 digits			93 <sup>rd</sup> percentile			
Digit span forward + back <sup>c</sup>									
			Score = 24			91 <sup>st</sup> -97 <sup>th</sup> percentile			
2-back	30	87%	96.4%	4.0%	18	2.06	17	.03	2.35

Note. DL = DL's performance, either span or percentage of correct responses.

<sup>a</sup>No statistical comparison when DL's score was identical with, or numerically higher than, the control group. <sup>b</sup>Control data for these span tasks are taken from Gvion and Friedmann (2012). <sup>c</sup>These span tasks and their norms are from WAIS-III (Wechsler, 1997).

## Verbal Memory

**Short-term memory and working memory** — We examined the two components of verbal-phonological short-term memory: the **phonological input buffer**, the short-term store that maintains auditory verbal information in memory during comprehension; and the **phonological output buffer**, the short-term store that maintains phonological elements during speech production (Butterworth, 1989, 1992; Dell, 1986, 1988; Franklin, Buerk, & Howard, 2002; Friedmann, Biran, & Dotan, 2013; Friedmann & Gvion, 2002; Garrett, 1976, 1992; Gvion & Friedmann, 2012; Kempen & Huijbers, 1983; Levelt, 1989, 1992; Martin, Shelton, & Yaffee, 1994; Monsell, 1987; Nickels, 1997; Nickels, Howard, & Best, 1997; Patterson & Shewell, 1987; Shallice, Rumiati, & Zadini, 2000; Shallice & Warrington, 1977). We also examined DL's working memory – the mechanisms that maintain mental representations available for use in thought and action – i.e., situations where the information is both maintained and being used actively (Oberauer et al., 2018).

### Span tasks

The **phonological input buffer** was examined using tasks that required memorizing an auditory input sequence but did not require producing it verbally: (1) Digit matching span – DL heard pairs of digit sequences in increasing lengths, and judged whether the two sequences in each pair have the same digit order (e.g., 3-7; 3-7) or not (3-7; 7-3). (2) Word matching span – same, but with words rather than digits. To examine the **phonological output buffer**, we used tasks that required both memorizing the auditory input and producing it verbally: (1) reading aloud 40 nonwords (from the TILTAN battery, [Friedmann & Gvion, 2003a](#)). (2) Repeating 48 nonwords, some of which were long, and phonologically or morphologically complex (from the BLIP battery, [Friedmann, 2003](#)). (3) Serial recall (span) – repeating sequences of digits, words, or nonwords in increasing lengths (from the FriGvi battery, [Friedmann & Gvion, 2002](#); [Gvion & Friedmann, 2012](#); in the words and non-words tasks, the items in each sequence were phonologically and semantically dissimilar). DL's performance in all these tasks was good ([Table 2](#)), indicating good verbal-phonological short-term memory – i.e., intact phonological input buffer and phonological output buffer.

Working memory was examined using a **backward span** task (from WAIS-III; [Wechsler, 1997](#)): DL heard sequences of digits in increasing length, and repeated each sequence in reverse order. As control, she also performed the WAIS-III forward span task (same, without reversal). She performed well in both tasks ([Table 2](#)).

### 2-back

Another task we used to examine working memory is the **2-back** task (from the FriGvi battery, [Friedmann & Gvion, 2002, 2003b](#)): DL heard 99 one-syllable animal names, one item per second, and for each item she decided whether it was identical with the item that was 2 places back in the list (e.g., *cat* – *donkey* – *cat*; this was the case for 30 items). DL's performance was worse than the control group ([Table 2](#)).

Note that DL showed poor performance in the 2-back task, but good performance in the backward span task. This discrepancy between the tasks indicates that, although both tasks are typically considered as indexing working memory, they require somewhat-different cognitive abilities. We propose that the crucial difference between the two tasks is that the 2-back task, but not the backward span task, induces interference. The 2-back task requires attending a target item (in our task, the item that appeared 2 positions before in the list), and ignoring temporarily-irrelevant items (i.e., the  $n-1$  item). Crucially, the classification of each item as relevant or irrelevant keeps changing. This is parallel to interference when memorizing multiplication facts, which requires attending to a relevant fact and ignoring temporarily-irrelevant facts, while the classification of each fact as relevant or irrelevant keeps changing. Under this interpretation, DL has normal working memory capacity (demonstrated by her backward span), and her difficulty in the 2-back task resulted from a very specific deficit – hypersensitivity to interference. DL's difficulty in the 2-back task may resemble the difficulty of DB ([De Visscher & Noël, 2013](#)) in the “recent probes” task, in which interference was induced by an item from the previous trial.

**Long-term verbal memory** — DL also performed two memorization tasks that examined her verbal long-term memory. In the first task she memorized a list of arbitrary words, and in the second task – a short story.

### *Memorizing a list of words*

Rey AVLT test, Hebrew version (Vakil & Blachstein, 1997). The task included 10 sub-tests. The first 5 sub-tests were free recall of the same list of 15 nouns. The 6<sup>th</sup> sub-test was a free recall of a new list of 15 words, and in the 7<sup>th</sup> sub-test DL heard nothing and recalled the first list. The two latter sub-tests examine possible effects of interference between the two lists. The 8<sup>th</sup> sub-test, performed after a 20-minute retention interval that included no verbal tasks, required recalling list #1 again (without hearing it). The 9<sup>th</sup> sub-test required recognizing the 15 words of list #1 in a list of 50 nouns (the distracters being semantic, phonological, and list #2). Finally, DL was asked to sort the shuffled 15 words of list #1 to their original order. She performed well in all sub-tests (z scores: 0.57, 1.03, 0.25, 1.44, -0.05, -0.68, 0.56, 1.42, 0.73, 0.93), demonstrating good verbal short-term and long-term memory. We also examined 2 additional measures of the Rey AVLT test that specifically consider possible effects of interference. One measure, which aims to tap the effect of proactive interference, is the difference between the 1<sup>st</sup> and the 6<sup>th</sup> sub-tests (i.e., how well was list#2 memorized versus how well list#1 was first memorized). DL's score was slightly low on this measure ( $z = -1.11$ , one-tailed  $p = .13$ ), suggesting perhaps some sensitivity to proactive interference. A second measure, which aims to tap retroactive interference, is the difference between the 5<sup>th</sup> and 7<sup>th</sup> sub-test (i.e., the degradation in list#1 following the memorization of list #2). DL's score in this measure was good ( $z = 0.77$ ).

### *Memorizing a short story (Cohen, 1997)*

The experimenter read aloud two short stories (about 100 words each), one after another. DL repeated each story immediately after its presentation, and again after 30 minutes (during which other tasks were administered). Her performance was on the 34<sup>th</sup>-40<sup>th</sup> percentile both in immediate recall and in delayed recall. Again, this result indicates good short-term and long-term verbal memory.

### **Interim Summary**

DL performed well in the tasks that examined reading, lexical retrieval, verbal memory, and symbolic number processing. Thus, her difficulty in memorizing multiplication facts does not originate in a general deficit of language, number processing, or memory. These results replicate previously-reported dissociations between good memory functions and impaired knowledge of arithmetic facts (Butterworth, Cipolotti, & Warrington, 1996; Kaufmann, 2002).

DL had difficulty in only one memory task – the 2-back task. We proposed that this difficulty actually originated in her hypersensitivity to interference, because the 2-back task may induce a high degree of interference.

### **Sensitivity to Interference**

To examine DL's sensitivity to interference, we adapted to Hebrew De Visscher and Noël's (2013) "first name – surname – country" memorization task. The task requires memorizing two list of verbal, non-numeric facts: one list with high between-item similarity, and another list with low between-item similarity. If DL has hypersensitivity to interference, she should perform better on memorizing the low-similarity list than on memorizing the high-similarity list. This was the pattern exhibited by DB in De Visscher and Noël (2013), and as we shall now see – also by DL.

## Method

The task included a list of 12 fictitious person names (first name + surname) and a country in Africa or Asia where each person allegedly lived. Unknown to DL, the 12 names were the mixture of two lists with 6 names in each: in the low-similarity list, each first name and each surname appeared only once. In the high-similarity list, there were only 3 first names and 3 surnames, and each repeated twice to create 6 different combinations. DL memorized the 12-item list in 5 successive and identical learning stages, each of which was administered as follows. First, the experimenter said aloud each item (name-surname-country) and DL repeated it. After completing the list of 12 items, the experimenter said each name-surname (in random order), DL answered where that person lived, and the experimenter corrected her errors. The 5 learning stages were followed by a final test stage: DL was presented with 24 name-surname-country combinations, and judged whether each combination was correct or not. In this final test, each name-surname appeared twice – once with the correct country, and once with the country of one of the 5 other persons in the same list. Both during learning and during testing, two subsequent items never had the same first name, surname, or country. DL's performance in this task was compared with 24 age-matched control participants (mean age = 40;3,  $SD = 5;2$ , range = 31;6 to 48;5) with no reported cognitive deficits and with a mean digit span of 7.17 ( $SD = 1.19$ ). One additional control participant was excluded due to outlier performance (chance level) in the final test. The detailed results of this task (DL and controls) are available in Supplementary Online Material.

## Results

Notably, in the final test (Table 3) even the control group was affected by the similarity between list items: their performance in low-similarity items was marginally better than in high-similarity items, paired  $t(23) = 1.64$ , one-tailed  $p = .06$ , Cohen's  $d = 0.41$ . This effect could also be observed during learning – e.g., in the last learning stage they recalled 4.21 out of 6 low-similarity items ( $SD = 1.72$ ), but only 2.79 out of 6 high-similarity items ( $SD = 1.28$ ; paired  $t(23) = 4.30$ , one-tailed  $p = .0001$ ,  $d = 0.92$ ). These results agree with previous findings of similarity-induced interference in normal population (Corman & Wickens, 1968; Hall, 1971; Mark-Zigdon & Katzoff, 2015; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012; Oppenheim, Dell, & Schwartz, 2010; Posner & Konick, 1966; Runquist, 1970, 1971).

Table 3

*Number of Correct Responses (out of 12) in the Verbal Memorization Task (Name – Surname – Country). DL performed poorly in the high-similarity items, which are especially sensitive to interference, but she performed well in the low-similarity items*

Similarity between items	DL	Control group		DL vs. controls (one-tailed $p$ ) <sup>a</sup>			
		<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>z<sub>cc</sub></i>
Low	11	10.32	1.44	–	–	–	–
High	7	9.88	1.26	2.24	23	.02	2.29

*Note.* DL = DL's performance, number of correct responses (out of 12) in the Verbal Memorization Task (Name – Surname – Country).

<sup>a</sup>No statistical comparison for low-similarity items, because DL's score was numerically higher than the control group.

As for DL, she showed a dramatic effect of similarity in the final test: she performed almost at ceiling on the low-similarity list, having only a single error – fewer errors than the control group, and not significantly more than zero errors (Fisher's one-tailed  $p = .50$ ). Conversely, her performance was poor on high-similarity items – significantly worse than the control group, not significantly different from the chance level of 50% (Fisher's one-tailed  $p = .50$ ), and significantly worse than her own performance in the low-similarity items (using the paired

bootstrap method described in General Method, one-tailed  $p = .01$ ,  $z = 1.83$ ). This pattern meets the criteria for classical dissociation (Crawford, Garthwaite, & Gray, 2003). Crucially, increasing the similarity level (low-similarity list versus high-similarity list) disrupted DL's performance significantly more than it disrupted the control group's performance (dissociation analysis of Crawford, Garthwaite, & Porter, 2010:  $t(23) = 2.15$ , one-tailed  $p = .02$ ). These results clearly show that DL was sensitive to the level of item similarity significantly more than the control group. Namely, she had hypersensitivity to verbal interference.

## Summary of the Assessment Results

DL's difficulty in solving multiplication facts is best explained as hypersensitivity to verbal interference: she demonstrated this hypersensitivity also in a memorization task that did not involve numbers. In contrast, she showed good general memory, language, and symbolic number processing abilities.

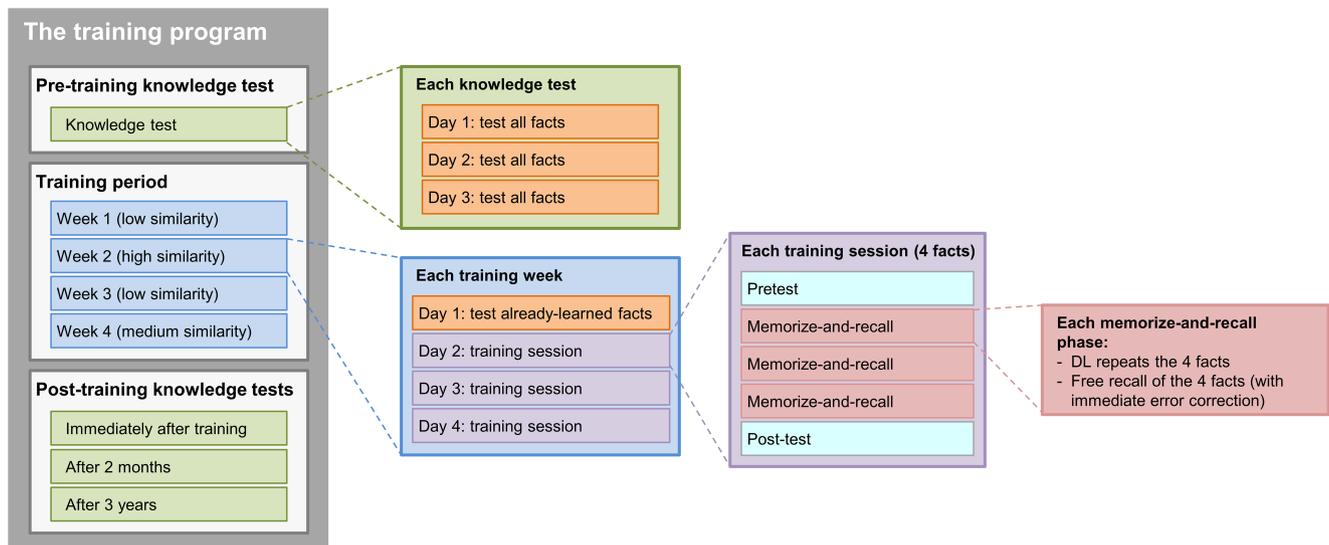
## Multiplication Facts Training

We now turn to the main aim of this study – to examine whether our training method would allow DL to learn the multiplication table in spite of her hypersensitivity to interference.

### Method

The training program was structured as a pre-training test, a training period, and several post-training tests (left column in Figure 1). The training was done on the 16 multiplication facts with the lowest pre-training scores, which were grouped into 4 sets with 4 facts in each. Each set of facts was trained during one week, and only during this week. After the 4-week training period, a post-training test evaluated DL's knowledge of all multiplication facts. Another test was run after 2 months, during which DL received no training. Throughout this 3-month period, DL was asked not to rehearse multiplication on her spare time, and she reported to have followed this instruction. All training and test sessions were performed orally over the telephone, while DL was in a quiet room in her home. In all training and testing, the larger operand always appeared first.

To allow examining the effect of interference, the four sets of trained facts were constructed, unknown to DL, to have different degrees of within-set interference: there were three low-interference sets and one high-interference set. The degree of interference was manipulated by controlling the degree of similarity between the facts in a given set. We predicted that DL would show better memorization of the low-interference sets than of the high-interference set. As we shall see below, this prediction was confirmed, which means that at the end of the study DL knew some multiplication facts but not others. Thus, after completing the study (including the 2-month follow-up test), we taught DL the remaining facts properly (in the low-interference mode), so that by the time she left our lab she knew the multiplication table fully. An additional test of DL's knowledge was run 3 years later.



*Figure 1.* DL's training program was structured as a pre-training knowledge test, a training period, and post-training knowledge tests in three different time points. Each knowledge test (green) examined DL's knowledge of all multiplication facts three times, in three different days of one week. Training was done in four weeks; each week (blue), in which four facts were trained, comprised of four sessions, held in different days: first, a session for testing DL's knowledge of the facts learned during the previous weeks (orange); then, three training sessions for learning the new set of facts (purple). Each training session started and ended with a short knowledge test; in between, there were three memorization-and-recall cycles (red).

### Grouping the Trained Facts Into Sets

The 16 trained facts were grouped into 4 sets, with 4 facts per set. The sets differed from each other in the level of within-set similarity, which was computed using De Visscher and Noël's (2014b) method: first, the similarity between each two multiplication facts was defined as the number of digit pairs that appeared in both facts, irrespectively of the digits' position in the fact and of the relative order of the two digits. For example, the facts  $8 \times 7 = 56$  and  $8 \times 3 = 24$  have no common digit pair (only the digit 8 appears in both) so their similarity is 0. The facts  $3 \times 4 = 12$  and  $3 \times 7 = 21$  have three common digit pairs (1-2, 2-3, and 1-3) so their similarity is 3. Then, the similarity index for a set of 4 facts was computed by summing the pairwise similarities of all 6 fact pairs in the set. Below, in Appendix A, we consider alternative methods to compute similarity and their fit to the observed results.

Of the four sets of facts, one set had high similarity ( $7 \times 4 = 28$ ,  $7 \times 6 = 42$ ,  $8 \times 4 = 32$ ,  $9 \times 4 = 36$ , similarity=9). The three other sets had lower similarities ( $4 \times 4 = 16$ ,  $8 \times 3 = 24$ ,  $8 \times 7 = 56$ ,  $5 \times 3 = 15$ , similarity=0;  $8 \times 8 = 64$ ,  $9 \times 7 = 63$ ,  $6 \times 2 = 12$ ,  $8 \times 6 = 48$ , similarity=3;  $9 \times 6 = 54$ ,  $6 \times 5 = 30$ ,  $8 \times 5 = 40$ ,  $7 \times 5 = 35$ , similarity=4).

### The Training Program

Each set of 4 facts was trained during one week, in four 5-minute sessions held in four separate days. Each week started with a session that tested DL's knowledge of the facts that she learned during the previous weeks, and continued with 3 identical training sessions (Figure 1). Each session took about 5 minutes. The high-similarity set was trained on the second week.

**Training sessions.** The session started with a pretest: the experimenter said each fact and DL said the answer. After responding to all 4 facts, her errors were corrected. Next, 3 memorization-and-recall phases were

done. In each phase, the experimenter said the 4 facts (exercise and result, e.g., "four times five, twenty"), presented in the same order in each session, and DL repeated each fact immediately after it was said; then DL recalled the 4 facts in free recall. Errors, "I don't know" responses, and omissions of facts were corrected immediately. A post-test, administered like the pretest, completed the session. The order of presenting the facts in each session, and DL's full responses, are available as Supplementary Online Material.

**Testing during training.** In the first session in each week (except the first week), we tested DL's knowledge of all the facts she learned since the beginning of the training program. Namely, 4 facts were tested in the beginning of the 2<sup>nd</sup> week, and all 16 facts were tested in the beginning of the 5<sup>th</sup> week. In each of these sessions, each fact was presented 3 times in a pseudo-random order, such that the same fact never appeared twice in a row, and the fact  $a \times b$  was never followed by the facts  $a \times (b \pm 1)$  or  $(a \pm 1) \times b$ . Errors were not corrected, and there was no teaching during these test sessions. DL's full list of responses in these tests is available as Supplementary Online Material.

### Testing Before and After the Training

DL's knowledge of the multiplication facts was tested in 4 time points (hereby, "testing times"): before training, immediately after training, two months after the training ended, and 3 years after the study ended. Each testing time consisted of 3 testing sessions, administered in 3 days of a single week. In each of these testing sessions, DL was asked to solve the 55 multiplication facts. Reaction times were defined as the delay between the time when the experimenter finished asking the question and the time when DL started saying the result. This delay was measured by inspecting the recordings with an audio-processing software.

Three kinds of responses were classified as errors: (1) Incorrect responses, even if preceded or followed by a correct response. (2) "I don't know" responses. (3) Extremely slow responses, which suggest that DL used calculation rather than retrieval: reaction times that exceeded the 75<sup>th</sup> percentile of the correct-response trials by more than 150% the inter-quartile range. To compute the outlier threshold, items that were classified as errors by one of the two other criteria were not included. The outlier threshold was computed per testing time, but essentially the same results were obtained when using a common threshold for all testing times.

In each testing time, DL was tested 3 times on all facts, yielding an accuracy score between 0 and 3 for each fact. The 16 facts with lowest pre-training scores were selected for training (10, 3, and 3 facts with score = 0, 1, and 2, respectively).

On top of these four testing times, we also examined DL's performance in the testing sessions that were administered during the training period, in the first session in each week. In each of these sessions, DL was tested on the facts that she already learned (i.e., on 4 facts in the beginning of the second week, and on 16 facts in the beginning of the fifth week). From each session we considered only the facts learned in the previous week. Below, we refer to this data as "testing during training".<sup>iii</sup>

## Results

### Effectiveness of Training: Performance Over all Facts

In the rule-based facts ( $N \times 0$  and  $N \times 1$ ), DL had merely a single error in the pre-training test and in the 3-year follow-up, and no error in the other testing times. Table 4 shows her error rates in the non-rule facts. Outlier reaction times of the non-rule facts were classified as errors, as explained in the previous section ("Testing Be-

fore and After the Training”): responses slower than 2408 ms in the pre-training test, 3265 ms in the post-training test, 4010 ms in the 2-month follow-up test, and 2380 ms in the 3-year follow-up test. To examine the effect of training, we compared the pre-training performance with each of the other testing times using the paired bootstrapping method described in General Method, based on DL’s pre-training and post-training scores (a 0-3 score for each fact).

Table 4

*DL’s Percentages of Correct Responses Before, During, and After Training in the Non-Rule Facts. Her knowledge of trained facts significantly increased following the training, and this improvement persisted two months later, in the follow-up test*

Test	Before training	During training		After training		2-month follow-up		3-year follow up	
	% correct	% correct	$z_{bs}$	% correct	$z_{bs}$	% correct	$z_{bs}$	% correct	$z_{bs}$
All 36 facts	56			75***	2.87	81***	3.72	62	
16 trained facts	21	79***	3.41	56***	2.84	65***	3.18	35**	1.94
20 untrained facts	83			90		95**		83	

Note.  $z_{bs}$  = Bootstrap effect size.

\*\* $p < .01$  (one-tailed). \*\*\* $p < .001$  (one-tailed). Comparison with pre-training test.

DL’s performance in the non-rule facts (Table 4) significantly improved from the pre-training test to the post-training test, i.e., the training was effective. As predicted, this overall improvement was driven by a significant improvement in the trained facts, with no significant improvement in the untrained facts<sup>iv</sup>. Strikingly, DL’s relatively short training program has succeeded where years of schooling did not: it gave rise to an improvement in the trained facts. This improvement persisted even after 3 years during which she received no additional training (Table 4), and – according to her – did not use the multiplication table very often. Still, unsurprisingly, her knowledge of the trained facts after 3 years was not as good as at the end of the training program (comparing the two using the bootstrapping method described above, one-tailed  $p = .03$ ,  $z = 1.69$ ).

### Effect of Within-Set Interference on the Post-Training Knowledge

Our main question was whether we can detect differences in the training effect between the facts that were trained together with highly-similar facts and the facts that were trained with dissimilar facts. Figure 2 presents DL’s detailed performance for each exercise in each of 4 testing times – before training, during training (on the testing sessions in the beginning of each week), after training, and in the 2-month follow-up<sup>v</sup>. The table clearly shows an effect of similarity: both during training and in the post-training tests, accuracy was better in the sets with lower within-set similarity than in the high-similarity set. To examine whether this effect was significant we considered, for each testing time, the within-set similarity values corresponding with each of DL’s answer attempts. We classified these values into two groups according to her answer (correct or incorrect), and compared the two groups using Mann-Whitney  $U$  test (the effect size  $r$  was computed according to Fritz, Morris, & Richler, 2012). The correct answers had significantly lower similarity values than the incorrect answers, confirming that lower within-set similarity in the training period improved the post-training accuracy (during training:  $U = 94$ , one-tailed  $p = .006$ ,  $r = .36$ ; after training:  $U = 217.5$ , one-tailed  $p = .08$ ,  $r = .20$ ; two-month follow-up:  $U = 161.5$ , one-tailed  $p = .01$ ,  $r = .33$ ; i.e., a medium effect size during training and in the 2-month follow-up). To eliminate any possible effect of prior knowledge, we also ran the same analysis only on the 9 exercises with pre-training score = 0. The results were essentially the same: for the tests administered during training, this

analysis showed a slightly stronger effect than in the previous analysis ( $U = 35.5$ , one-tailed  $p = .01$ ,  $r = .43$ ). For the tests administered after training, the analysis now showed a slightly weaker effect than before, which was marginally significant (immediately after training:  $U = 58$ , one-tailed  $p = .08$ ,  $r = .27$ ; two-month follow-up:  $U = 64$ , one-tailed  $p = .09$ ,  $r = .26$ ).

These results were not an artifact of problem size (better knowledge of facts with smaller operands, Zbrodoff & Logan, 2005). First, a fact's average operand size did not correlate with the within-set similarity level ( $r = .20$ ,  $p = .46$ ). Second, when directly comparing operand size and similarity as explanations to DL's performance during and after training, similarity stands out as a genuine predictor of accuracy. This was examined by submitting DL's accuracy scores to a logistic regression with two predictors: the average operand size and the within-set similarity level – 1, 2, or 3 (weeks 3 and 4, whose similarities were nearly identical, were both assigned similarity level 2). Both during training and in the 2-month follow-up test, the within-set similarity had a significant effect on accuracy (during training:  $z = 2.59$ , one-tailed  $p = .005$ ; follow-up:  $z = 2.23$ , one-tailed  $p = .01$ ), whereas the problem size had a smaller or non-significant effect (during training:  $z = 0.95$ , one-tailed  $p = .17$ ; follow-up:  $z = 1.73$ , one-tailed  $p = .04$ ). Namely, the effect of similarity could not be reduced to a problem size effect. In the post-training test, however, the regression showed a significant effect only for problem size ( $z = 2.46$ , one-tailed  $p = .007$ ), with no significant effect of similarity ( $z = 0.76$ , one-tailed  $p = .22$ ).

A second analysis ignored the specific similarity values, and just compared the per-fact accuracy after training (correct or incorrect for each answer attempt) between the high-similarity set (H) and the three lower-similarity sets (L). The comparison used the unpaired bootstrapping method described in General Method, based on the 0-3 score of each fact (the random distribution for  $H_0$  was computed by considering all  $\binom{16}{4}$  possible classifications of the 16 facts into arbitrary "H" and "L" groups containing 4 and 12 facts respectively). The effect of similarity was significant in the tests administered during training (H: 50%, L: 89%, one-tailed  $p = .003$ ,  $z = 1.85$ ), in the post-training test (H: 42%, L: 61%,  $p = .03$ ,  $z = 0.81$ ), and in the two-month follow-up test (H: 17%, L: 81%,  $p < .001$ ,  $z = 2.81$ ). These results survived the exclusion of the three facts with the highest pre-training scores (during training: one-tailed  $p = .003$ ,  $z = 1.47$ ; post-training:  $p = .02$ ,  $z = 0.55$ ; two-month follow-up:  $p < .001$ ,  $z = 2.42$ ).

A third analysis examined how quickly DL learned each fact until reaching ceiling performance. Each fact's *learning duration* was defined as the last answer attempt, out of the fact's 12 answer attempts (right columns in Figure 2), in which DL made an error (here we did not code slow answers as errors, because most answer attempts during training were made as free recall, in which the response time is undefined). The 4 facts in the high-similarity set were the slowest to be learned – they had the longest learning durations ( $p = 1/\binom{16}{4} < .001$ ).

Taken together, these results form a clear picture: DL's learning of the multiplication facts was strongly affected by similarity. The effect of similarity was clearly observed during the 4-week training period, and critically – also in the test administered two months after DL's training has ended. In the test administered immediately after the 4-week training period, the effect of similarity was observed only in one of the two analyses. The weak similarity effect immediately after the 4-week training period can be explained by the fact that in this test, the between-fact variance in the time elapsed since each fact was learned was the highest: DL had learned the facts of set #4 only few days before the post-training test, but she had learned the facts of set #1 almost 4 weeks before the post-training test. The idea of time-elapsed-since-learning as a confounding factor is supported by comparing the results of the immediately-after-training test with those of the two other post-training tests (during train-

ing, 2-month follow-up): in the immediately-after-training test, DL performed more poorly for facts that were learned more recently. Whether this effect is random or has a genuine cognitive origin remains an open question, as the amount of data in the present study is insufficient to evaluate the effect reliably.

	Within-set similarity		Score				Answers During Training (each cell is one answer attempt)													
			Before training	During training	After training	Follow-up	Session 1			Session 2			Session 3							
Week 1	Low (S = 0)	4 × 4 = 16	0	3	3	3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		8 × 3 = 24	0	3	2	2	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	8 × 7 = 56	0	2	0	0	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	5 × 3 = 15	2	3	3	3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	Total	2	11	8	8															
Week 2	High (S = 9)	7 × 4 = 28	0	0	2	1	✓	✓	✓	✓	✗	✓	✗	✗	✓	✓	✗	✓	✓	✓
		7 × 6 = 42	1	2	3	0	✗	✗	✓	✗	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓
	8 × 4 = 32	0	0	0	0	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	
	9 × 4 = 36	0	3	0	1	✗	✗	✗	✗	✓	✗	✗	✓	✗	✓	✓	✓	✓	✓	
	Total	1	5	5	2															
Week 3	Low (S = 3)	8 × 8 = 64	0	3	2	2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9 × 7 = 63	0	0	0	2	✗	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	6 × 2 = 12	1	3	3	3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	8 × 6 = 48	2	3	3	3	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	
	Total	3	9	8	10															
Week 4	Low (S = 4)	9 × 6 = 54	0	2	1	2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		6 × 5 = 30	1	2	1	3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	8 × 5 = 40	1	3	3	3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	7 × 5 = 35	2	3	1	3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	Total	4	10	6	11															

Figure 2. DL's performance before, during, and after training. During training, DL's score in the low-similarity set was better than in the high-similarity set. This difference persisted when she was tested 2 months later. The within-set similarity index (S) was computed as explained in the "Grouping the Trained Facts Into Sets" section. The "Score" columns show the score for each trained fact in each testing time (scale: 0-3 per fact, 0-12 per set; red/green shading indicates below/above 50% correct). The "Answers During Training" columns show DL's accuracy in each answer attempt during the training sessions.

The effect of similarity did not go unnoticed by DL herself: during the 2<sup>nd</sup> week of training, when she learned the high-similarity set, she commented more than once that "it is hard for me to learn these exercises because of all these 4's that repeat over and over again" – an accurate description of her hypersensitivity to interference.

**Pre-Training Scores**

If DL's learning is affected by the similarity between facts, this should also be reflected in her performance in the pre-training test, because even before we started the training program, DL's knowledge of specific arithmetic facts was presumably affected by similarity-induced interference (De Visscher & Noël, 2014b). In particular,

we predicted that she would show lower pre-training knowledge for multiplication facts that have higher similarity with the rest of the multiplication table.

To examine this prediction, we considered each of DL's answer attempts in the pre-training test for all 36 non-rule facts. For each answer attempt, we computed the fact's similarity with the rest of the multiplication table (hereby denoted *Sim*) as the sum of that fact's similarities with all other multiplication facts between 2×2 and 9×9 (the similarity between each pair of facts was defined as explained in the "Grouping the Trained Facts Into Sets" section). We predicted that DL's pre-training knowledge would be better for facts with lower fact-table similarity (lower *Sim*). Namely, the *Sim* values should be larger for DL's incorrect pre-training answers than for her correct answers. To assess whether this was the case, we used the unpaired bootstrapping method described in General Method. The difference between *Sim* of correct answers (15.25) and that of incorrect answers (18.19) was significant (one-tailed  $p = .03$ ,  $z = 1.83$ ). Essentially the same results were obtained when the fact-table similarity computation excluded the tie problems (*Sim* = 13.78 versus 16.21, one-tailed  $p < .05$ ,  $z = 1.72$ ). Namely, as predicted, DL's pre-training knowledge was better for multiplication facts that were less similar to the rest of the multiplication table.

## Discussion

### Hypersensitivity to Interference as a Source for Difficulty in Memorizing Arithmetic Facts

We reported the case of DL, a 40-year-old woman with severe difficulties in memorizing the multiplication table. DL had several spared memory functions: her short-term memory spans were in the normal range, she performed well in nonword reading and repetition, and she showed good ability to remember arbitrary lists of words and the details of a story for several tens of minutes. These results indicate good verbal short-term and long-term memory abilities, i.e., DL's difficulties in memorizing multiplication facts did not stem from a general deficit in verbal memory. This pattern is consistent with the finding of a double dissociation between memory capacity and knowledge of arithmetic facts (Butterworth et al., 1996; Kaufmann, 2002), and with the finding that short-term memory capacity and arithmetic fact knowledge do not correlate (Temple & Sherwood, 2002).

In contrast, DL performed poorly in a task that taps hypersensitivity to interference: when asked to memorize verbal non-numeric items, she performed poorly only in items that were similar to each other. Moreover, among the tasks that assessed working memory, DL's performance was poor only in the task that induced interference by repeating the same words over and over again (the 2-back task). Thus, her difficulty in multiplication is best explained as resulting from hypersensitivity to interference.

Our study aimed to confirm this conclusion by showing, for the first time, causal evidence for relation between hypersensitivity to interference and difficulty in multiplication facts. Our second goal was clinical – to develop a method to help individuals who are hypersensitive to interference to learn the multiplication table. To accomplish these aims, we devised a training method that manipulated the degree of interference. The method was clearly successful: DL managed to memorize multiplication facts as long as in a given week, she only had to learn facts that were relatively dissimilar from each other. In this condition, her learning was virtually immediate: in the set with lowest similarity, she reached perfect performance after merely two (!) exposures to each fact.

This good memorization of multiplication facts was exhibited during the training sessions, and even two months after DL's training has ended. In marked contrast, she had difficulty in the set with high similarity between facts: she made many errors while learning, her scores in the weekly tests during training were hardly any better than the pre-training score, and this small improvement virtually disappeared two months later. We showed that this effect of similarity could not be explained as an artifact of problem size.

These results extend the findings of [De Visscher and Noël \(2013\)](#) in two ways. First, the hypersensitivity to interference of the woman they reported, DB, was manifested mostly in slow retrieval of multiplication facts, whereas DL showed not only slow RTs but actually erred in almost half of the multiplication facts (even when ignoring the RT outliers, DL's error rate in non-rule facts before the training program was 39%). Second, whereas De Visscher and Noël's evidence for interference as the source of multiplication difficulty was correlational, here we showed evidence for a causal relation: manipulating the amount of similarity-induced interference affected the memorization of multiplication facts.

## The Cognitive Mechanisms Underlying Sensitivity to Interference

Our findings clearly show that DL's difficulty in multiplication was the result of hypersensitivity to interference. Still, to understand the exact origin of interference, we need to understand the cognitive architecture that underlies the representation of arithmetic facts, and the process that is sensitive to interference.

### The Processing Stage Sensitive to Interference

According to Campbell's network interference model, interference arises because one multiplication fact is associated with another, so they activate overlapping representations ([Campbell, 1987, 1995](#); [Oberauer, 2009](#)). When the facts are similar to each other, irrelevant associations between them are strengthened. Furthermore, processing different facts in temporal proximity may strengthen the irrelevant associations between these facts. In our experiment, manipulating similarity and temporal proximity during learning affected how well DL learned the facts. Importantly, the effect of similarity could reliably be attributed to processes that occur during learning time, because similarity and temporal proximity were not manipulated when DL's knowledge was tested (but note that the learning sessions included not only encoding and storage but also retrieval). Additionally, [Campbell \(1987\)](#) showed that arithmetic performance was affected by manipulating the degree of interference during retrieval, for arithmetic facts that the participants already knew beforehand. It therefore seems that interference affects both learning-time and retrieval-time processes. Indeed, other studies too showed that high levels of interference may take an effect in different processing stages ([Bartko, Cowell, Winters, Bussey, & Saksida, 2010](#); [Farrell, 2006](#); [Fernandes & Moscovitch, 2000](#); [Kaufmann, Lochy, Drexler, & Semenza, 2004](#); [Lochy, Domahs, & Delazer, 2004](#); [Van Dyke & McElree, 2006](#); [Wixted, 2004](#)). Interference may disrupt the encoding and storage of data in memory while learning the facts ([Farrell & Lewandowsky, 2002](#); [Lewandowsky & Farrell, 2008](#)), and it may disrupt the retrieval stage ([Burgess & Hitch, 1999](#); [Campbell, 1987](#); [Henson, 1996, 1998](#)).

A more specific question concerns the locus of DL's deficit. Her hypersensitivity to interference may have resulted from an encoding/storage deficit in creating the network of associations in long-term memory – e.g., she may have been creating too strong irrelevant associations. Alternatively, her deficit may be impaired retrieval processes, which fail retrieving the arithmetic facts from an intact storage. Sadly, it seems that our data cannot arbitrate between these possibilities. Within the framework of the network interference model, retrieving the cor-

rect answer to a multiplication problem requires that the representation of the fact is distinctive enough from other facts. When this is not met, incorrect responses occur. This can happen because the storage of facts is corrupted such that the facts are not distinguishable sufficiently; but it can also happen because the retrieval process is impaired and requires higher distinctiveness for successful retrieval. Under either assumption, increasing the distinctiveness of facts (as our manipulation probably did) would increase the activation of the correct solution relative to incorrect solutions, and would therefore help achieving the required threshold of distinctiveness. In short, both possibilities (storage deficit and retrieval deficit) make similar predictions about DL's performance. Note, however, that from a clinical/intervention point of view, the picture emerging from our study is clear: a learning-time intervention can help overcoming similarity-induced interference.

### Interference and Spacing

De Visscher and Noël (2014b) described interference as an effect of previously-learned items on the ability to learn a new item. Consequently, they defined an “interference parameter” for each multiplication fact as its degree of similarity versus all previously-learned facts. This formulation accords with the definition of proactive interference in working memory (Bennett, 1975). However, in the present study, it is impossible to define a clear order of learning the facts, because several multiplication facts were learned simultaneously during each given week (this is probably the case also in real-life situations of learning the multiplication table). The interference here does not arise from the similarity with previously-learned items, but from the similarity with *simultaneously*-learned items.

Our data suggest that the critical methodological factor is to create a sufficient temporal gap between the learning of similar multiplication facts. Future studies may examine the size of the temporal gap required to avoid interference. This idea of temporal separation accords with other studies showing that learning is improved by increasing the temporal delay between learning sessions that contain potentially-interfering items (Friedman & Korman, 2016).

### The Type of Information Sensitive to Interference

The effect of similarity on the size of interference accords with the view of interference as arising from the amount of overlapping features between the items to be remembered (Oberauer & Lange, 2008). But what are these features? The representations sensitive to interference could be phonological (Baddeley, 1966, 1968; Farrell, 2006; Nelson et al., 1974; Runquist, 1970), semantic (Baddeley, 1966; Oppenheim et al., 2010), number-specific, or another representation.

In line with the possibility of phonological sensitivity-to-interference, the speed and accuracy of addition fact retrieval was shown to be affected by phonological similarity (Noël, Désert, Aubrun, & Seron, 2001). Further support to the role of phonological interference comes from studies of non-number words, which show that word memorization is affected by their phonological similarity to each other (Nelson et al., 1974; Pajak et al., 2016; Runquist, 1970). However, interpreting these findings as an explanation for difficulties in memorizing multiplication facts should be done with caution, because at least some phonological mechanisms treat words and numbers differently: e.g., the speech mechanisms handle words as sequences of phonemes, but numbers words as whole building blocks – in speech production (Bencini et al., 2011; Cohen, Verstichel, & Dehaene, 1997; Dotan & Friedmann, 2015, 2019; Shalev, Ophir, Gvion, Gil, & Friedmann, 2014), and apparently also in speech comprehension (Fischer-Baum, Mis, & Dial, 2018). Furthermore, the representation of multiplication facts in memory is apparently not purely phonological (Whalen, McCloskey, Lindemann, & Bouton, 2002).

In order to fully understand sensitivity to interference, and the effect of similarity, one must use a detailed cognitive model of arithmetic fact representation. Only a detailed model can specify the precise amount of representational overlap between some given facts, and the amount of interference between them. Consequently, each such model can predict how easy it should be to learn a given set of multiplication facts, so the model can be evaluated by comparing these predictions against the actual learnability of facts. In [Appendix A](#) we describe 7 detailed cognitive models of arithmetic fact representation, and we evaluate them based on their ability to predict several aspects of DL's performance. This analysis lends support to the models that assume symbolic representations (digits / number words) of arithmetic facts, rather than magnitude representation – in accord with several studies indicating that multiplication facts are stored in verbal format ([Dehaene, 1992](#); [Dehaene & Cohen, 1995](#); [Dehaene et al., 2003, 1999](#)). Future studies may elaborate further on the precise representation of arithmetic facts.

## Clinical and Pedagogical Implications

The clinical goal of this study was to examine whether a person can learn the multiplication table even when they have hypersensitivity to interference. Our results clearly indicate that they can: as long as we maintained a low level of interference, DL easily learned the multiplication facts. This is not trivial: conceivably, one could hypothesize that learning a sequence of facts such as  $7 \times 4$ ,  $8 \times 4$ , and  $9 \times 4$  would actually be easier – for example, it may more transparently lead to an addition-based strategy as scaffold for multiplication. The finding that such a set was harder to memorize, in spite of the opportunity for scaffold strategies, emphasizes even further the importance of similarity as a factor that determines the difficulty of memorization, at least for individuals with hypersensitivity to interference.

Our training method was effective for DL, a woman with hypersensitivity to interference, but its clinical implication may be most relevant for children who learn the multiplication table at school, many of whom may have normal sensitivity to interference. Will the same method be effective for all children, including children without hypersensitivity to interference? The findings of similarity-induced interference in [Campbell \(1987\)](#), whose participants were not screened for sensitivity to interference, suggest that the answer to this question is affirmative.

Additional support to our conclusions comes from another study that showed causal evidence for similarity-induced interference in typically-developing grade children ([Mark-Zigdon & Katzoff, 2015](#)). Similarly to our study, Mark-Zigdon and Katzoff examined how the memorization of multiplication facts is affected by manipulating the interference level. They taught a group of typically-developing 3<sup>rd</sup> grade children a set of 10 new multiplication facts, and showed that the children's memorization of these facts was disrupted if interference was induced by teaching a new set of multiplication facts immediately after the first set. Thus, like us, Mark-Zigdon and Katzoff showed that high-interference conditions disrupted memorization of multiplication facts. Still, each of the two studies highlights a slightly different aspect of interference: our study highlights the importance of low interference within a set of learned facts; Mark-Zigdon and Katzoff's study highlights the importance of avoiding interference from out-of-set facts. Together, the two studies support what we described in the Introduction as the two foundations of an interference-reducing training method: grouping dissimilar facts when teaching, and teaching different sets of facts with a sufficient temporal gap between them.

Our findings directly bear on the recommended practices for teaching the multiplication table. At least for individuals with hypersensitivity to interference, it seems that we should teach simultaneously dissimilar rather than

similar facts. This is almost the opposite of how multiplication is typically taught at school: very often, children learn the multiplication table in an ordered manner – first the products of 2, then the products of 3, etc. Although this ordered teaching method may have advantages, it implies that the children learn similar facts simultaneously, and this increases the degree of interference and may therefore create difficulty. Future studies may directly compare the traditional teaching method versus a low-similarity teaching method.

## Notes

i) In spite of her impaired knowledge of the multiplication facts, DL solved the three 2-digit multiplication exercises correctly. The reason was that 2 of the exercises involved only simple facts, which she knew, and for the 3<sup>rd</sup> exercise she used a bypass strategy ( $6 \times 7 = 6 \times 6 + 6$ ).

ii) These results are from the first pretest session described in the "Testing Before and After the Training" results section.

iii) Unlike the tests before and after training, here slow responses were not classified as errors, because we assumed that DL's speed may or may not have changed during the 4-week training period, so classifying RT outliers as errors might bias the results. Nevertheless, we verified that our conclusions were robust to this decision: the effect of similarity was observed also when classifying outlier RTs as errors, both when we computed the outlier RT threshold for each week separately and when we computed a single RT threshold for all 4 weeks.

iv) Conceivably, the increased performance in trained facts could have been explained as regression to the mean, because the pretest data presented here was also used to select the trained facts. However, we doubt that regression to the mean could convincingly account for tripling (!!) DL's accuracy. Regression to the mean also predicts a corresponding *decrease* in performance in the untrained facts, but no such decrease was observed. Crucially, regression to the mean cannot account for our main finding, reported in the next section: the effect of similarity on memorization.

v) The 3-year follow-up data was not included here. The reason is that after the experiment ended (i.e., after the 2-month follow-up), DL received a short additional training, which included – among others – the facts in the high-similarity set. This additional training, whose goal was purely clinical, cancels the difference between the experimental manipulations performed on the high-similarity and the low-similarity sets.

## Funding

This research was supported by INSERM; by CEA; by a grant from the Bettencourt-Schueller Foundation; by the Israel Science Foundation (grant no. 1066/14, Friedmann); by the Human Frontiers Science Program (RGP0057/201, Friedmann); and by the Australian Research Council Centre of Excellence for Cognition and its Disorders (CE110001021, <http://www.ccd.edu.au>). Dror Dotan is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship. The funders took no role in the research design, execution, analysis, interpretation, and reporting.

## Competing Interests

The authors have declared that no competing interests exist.

## Acknowledgments

We thank DL for her participation in this research and for her commitment to the long training procedure, and Ricardo Tarrasch, Maya Yachini, and Roy Luria for their advice. The research is a part of the doctoral dissertation of Dror Dotan in Tel Aviv University, under the supervision of Naama Friedmann and Stanislas Dehaene.

## Data Availability

For this study, a dataset is freely available (see the [Supplementary Materials](#) section).

## Supplementary Materials

DL's performance in the training program and in the sensitivity-to-interference assessment task: Raw results (for access, see Index of [Supplementary Materials](#) below).

**Index of Supplementary Materials**

Dotan, D., & Friedmann, N. (2019). *Supplementary materials to "Reducing interference improves the memorization of multiplication facts in case of hypersensitivity to interference"*. PsychOpen. <https://doi.org/10.23668/psycharchives.2674>

**References**

- Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology*, 18(4), 362-365. <https://doi.org/10.1080/14640746608400055>
- Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *The Quarterly Journal of Experimental Psychology*, 20(3), 249-264. <https://doi.org/10.1080/14640746808400159>
- Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4, 829-839. <https://doi.org/10.1038/nrn1201>
- Bartko, S. J., Cowell, R. A., Winters, B. D., Bussey, T. J., & Saksida, L. M. (2010). Heightened susceptibility to interference in an animal model of amnesia: Impairment in encoding, storage, retrieval – or all three? *Neuropsychologia*, 48(10), 2987-2997. <https://doi.org/10.1016/j.neuropsychologia.2010.06.007>
- Bencini, G. M. L., Pozzan, L., Bertella, L., Mori, I., Pignatti, R., Ceriani, F., & Semenza, C. (2011). When two and too don't go together: A selective phonological deficit sparing number words. *Cortex*, 47(9), 1052-1062. <https://doi.org/10.1016/j.cortex.2011.03.013>
- Bennett, R. W. (1975). Proactive interference in short-term memory: Fundamental forgetting processes. *Journal of Verbal Learning and Verbal Behavior*, 14(2), 123-144. [https://doi.org/10.1016/S0022-5371\(75\)80060-0](https://doi.org/10.1016/S0022-5371(75)80060-0)
- Biran, M., & Friedmann, N. (2004). *SHEMESH: Naming a hundred objects*. Tel Aviv, Israel: Tel Aviv University.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, 106(3), 551-581. <https://doi.org/10.1037/0033-295X.106.3.551>
- Butterworth, B. (1989). Lexical access in speech production. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 108–135). Cambridge, MA, USA: MIT Press.
- Butterworth, B. (1992). Disorders of phonological encoding. *Cognition*, 42(1–3), 261-286. [https://doi.org/10.1016/0010-0277\(92\)90045-J](https://doi.org/10.1016/0010-0277(92)90045-J)
- Butterworth, B., Cipolotti, L., & Warrington, E. K. (1996). Short-term memory impairment and arithmetical ability. *The Quarterly Journal of Experimental Psychology, Section A*, 49(1), 251-262. <https://doi.org/10.1080/713755603>
- Campbell, J. I. D. (1987). Network interference and mental multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(1), 109-123. <https://doi.org/10.1037/0278-7393.13.1.109>
- Campbell, J. I. D. (1995). Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Mathematical Cognition*, 1(2), 121-164.
- Campbell, J. I. D., & Graham, D. J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology*, 39(2), 338-366. <https://doi.org/10.1037/h0080065>

- Cohen, L., & Dehaene, S. (2000). Calculating without reading: Unsuspected residual abilities in pure alexia. *Cognitive Neuropsychology*, 17(6), 563-583. <https://doi.org/10.1080/02643290050110656>
- Cohen, L., Dehaene, S., Chochon, F., Lehéricy, S., & Naccache, L. (2000). Language and calculation within the parietal lobe: A combined cognitive, anatomical and fMRI study. *Neuropsychologia*, 38(10), 1426-1440. [https://doi.org/10.1016/S0028-3932\(00\)00038-5](https://doi.org/10.1016/S0028-3932(00)00038-5)
- Cohen, L., Verstichel, P., & Dehaene, S. (1997). Neologistic jargon sparing numbers: A category-specific phonological impairment. *Cognitive Neuropsychology*, 14(7), 1029-1061. <https://doi.org/10.1080/026432997381349>
- Cohen, M. J. (1997). *Children's memory scale*. San Antonio, TX, USA: The Psychological Corporation.
- Corman, C. D., & Wickens, D. D. (1968). Retroactive inhibition in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 7(1), 16-19. [https://doi.org/10.1016/S0022-5371\(68\)80157-4](https://doi.org/10.1016/S0022-5371(68)80157-4)
- Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia*, 40(8), 1196-1208. [https://doi.org/10.1016/S0028-3932\(01\)00224-X](https://doi.org/10.1016/S0028-3932(01)00224-X)
- Crawford, J. R., Garthwaite, P. H., & Gray, C. D. (2003). Wanted: Fully operational definitions of dissociations in single-case studies. *Cortex*, 39(2), 357-370. [https://doi.org/10.1016/S0010-9452\(08\)70117-5](https://doi.org/10.1016/S0010-9452(08)70117-5)
- Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, 27(3), 245-260. <https://doi.org/10.1080/02643294.2010.513967>
- Dagenbach, D., & McCloskey, M. (1992). The organization of arithmetic facts in memory: Evidence from a brain-damaged patient. *Brain and Cognition*, 20(2), 345-366. [https://doi.org/10.1016/0278-2626\(92\)90026-I](https://doi.org/10.1016/0278-2626(92)90026-I)
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44(1-2), 1-42. [https://doi.org/10.1016/0010-0277\(92\)90049-N](https://doi.org/10.1016/0010-0277(92)90049-N)
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY, USA: Oxford University Press.
- Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, 1, 83-120.
- Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, 33(2), 219-250. [https://doi.org/10.1016/S0010-9452\(08\)70002-9](https://doi.org/10.1016/S0010-9452(08)70002-9)
- Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, 20(3), 487-506. <https://doi.org/10.1080/02643290244000239>
- Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, 284(5416), 970-974. <https://doi.org/10.1126/science.284.5416.970>
- Delazer, M., & Benke, T. (1997). Arithmetic facts without meaning. *Cortex*, 33(4), 697-710. [https://doi.org/10.1016/S0010-9452\(08\)70727-5](https://doi.org/10.1016/S0010-9452(08)70727-5)

- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283-321. <https://doi.org/10.1037/0033-295X.93.3.283>
- Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, 27(2), 124-142. [https://doi.org/10.1016/0749-596X\(88\)90070-8](https://doi.org/10.1016/0749-596X(88)90070-8)
- De Visscher, A., & Noël, M. P. (2013). A case study of arithmetic facts dyscalculia caused by a hypersensitivity-to-interference in memory. *Cortex*, 49(1), 50-70. <https://doi.org/10.1016/j.cortex.2012.01.003>
- De Visscher, A., & Noël, M. P. (2014a). Arithmetic facts storage deficit: The hypersensitivity-to-interference in memory hypothesis. *Developmental Science*, 17(3), 434-442. <https://doi.org/10.1111/desc.12135>
- De Visscher, A., & Noël, M. P. (2014b). The detrimental effect of interference in multiplication facts storing: Typical development and individual differences. *Journal of Experimental Psychology: General*, 143(6), 2380-2400. <https://doi.org/10.1037/xge0000029>
- Dotan, D., & Friedmann, N. (2015). Steps towards understanding the phonological output buffer and its role in the production of numbers, morphemes, and function words. *Cortex*, 63, 317-351. <https://doi.org/10.1016/j.cortex.2014.08.014>
- Dotan, D., & Friedmann, N. (2019). Separate mechanisms for number reading and word reading: Evidence from selective impairments. *Cortex*, 114, 176-192. <https://doi.org/10.1016/j.cortex.2018.05.010>
- Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language*, 55(4), 587-600. <https://doi.org/10.1016/j.jml.2006.06.002>
- Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review*, 9(1), 59-79. <https://doi.org/10.3758/BF03196257>
- Fernandes, M. A., & Moscovitch, M. (2000). Divided attention and memory: Evidence of substantial interference effects at retrieval and encoding. *Journal of Experimental Psychology: General*, 129(2), 155-176. <https://doi.org/10.1037/0096-3445.129.2.155>
- Fischer-Baum, S., Mis, R., & Dial, H. (2018). Word deafness with preserved number word perception. *Cognitive Neuropsychology*, 35(8), 415-429. <https://doi.org/10.1080/02643294.2018.1515734>
- Franklin, S., Buerk, F., & Howard, D. (2002). Generalised improvement in speech production for a subject with reproduction conduction aphasia. *Aphasiology*, 16(10-11), 1087-1114. <https://doi.org/10.1080/02687030244000491>
- Friedman, J., & Korman, M. (2016). Offline optimization of the relative timing of movements in a sequence is blocked by retroactive behavioral interference. *Frontiers in Human Neuroscience*, 10, Article 623. <https://doi.org/10.3389/fnhum.2016.00623>
- Friedmann, N. (2003). *BLIP: Battery for assessment of phonological abilities*. Tel Aviv, Israel: Tel Aviv University.
- Friedmann, N., Biran, M., & Dotan, D. (2013). Lexical retrieval and breakdown in aphasia and developmental language impairment. In C. Boeckx & K. K. Grohmann (Eds.), *The Cambridge handbook of biolinguistics* (pp. 350-374). Cambridge, United Kingdom: Cambridge University Press.

- Friedmann, N., & Gvion, A. (2002). *FriGvi: Friedmann Gvion battery for assessment of phonological Working Memory*. Tel Aviv, Israel: Tel Aviv University.
- Friedmann, N., & Gvion, A. (2003a). *TILTAN: A test battery for dyslexias*. Tel Aviv, Israel: Tel Aviv University.
- Friedmann, N., & Gvion, A. (2003b). Sentence comprehension and working memory limitation in aphasia: A dissociation between semantic-syntactic and phonological reactivation. *Brain and Language*, *86*(1), 23-39.  
[https://doi.org/10.1016/S0093-934X\(02\)00530-8](https://doi.org/10.1016/S0093-934X(02)00530-8)
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*, 2-18. <https://doi.org/10.1037/a0024338>
- Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 231–256). Amsterdam, The Netherlands: North-Holland.
- Garrett, M. F. (1992). Disorders of lexical selection. *Cognition*, *42*(1–3), 143-180.  
[https://doi.org/10.1016/0010-0277\(92\)90042-G](https://doi.org/10.1016/0010-0277(92)90042-G)
- Girelli, L., Delazer, M., Semenza, C., & Denes, G. (1996). The representation of arithmetical facts: Evidence from two rehabilitation studies. *Cortex*, *32*(1), 49-66. [https://doi.org/10.1016/S0010-9452\(96\)80016-5](https://doi.org/10.1016/S0010-9452(96)80016-5)
- Graham, D. J., & Campbell, J. I. D. (1992). Network interference and number-fact retrieval: Evidence from children's alphaplication. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *46*(1), 65-91.  
<https://doi.org/10.1037/h0084310>
- Gvion, A., & Friedmann, N. (2012). Phonological short-term memory in conduction aphasia. *Aphasiology*, *26*(3–4), 579-614.  
<https://doi.org/10.1080/02687038.2011.643759>
- Hall, J. F. (1971). Formal intralist response similarity: Its role in paired-associate learning. *The American Journal of Psychology*, *84*(4), 521-528. <https://doi.org/10.2307/1421169>
- Henson, R. N. A. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology, Section A*, *49*(1), 80-115. <https://doi.org/10.1080/713755612>
- Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, *36*(2), 73-137.  
<https://doi.org/10.1006/cogp.1998.0685>
- Kaufmann, L. (2002). More evidence for the role of the central executive in retrieving arithmetic facts? A case study of severe developmental dyscalculia. *Journal of Clinical and Experimental Neuropsychology*, *24*(3), 302-310.  
<https://doi.org/10.1076/jcen.24.3.302.976>
- Kaufmann, L., Lochy, A., Drexler, A., & Semenza, C. (2004). Deficient arithmetic fact retrieval—Storage or access problem? A case study. *Neuropsychologia*, *42*(4), 482-496. <https://doi.org/10.1016/j.neuropsychologia.2003.09.004>
- Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, *14*(2), 185-209. [https://doi.org/10.1016/0010-0277\(83\)90029-X](https://doi.org/10.1016/0010-0277(83)90029-X)
- Lampl, Y., Eshel, Y., Gilad, R., & Sarova-Pinhas, I. (1994). Selective acalculia with sparing of the subtraction process in a patient with left parietotemporal hemorrhage. *Neurology*, *44*, 1759-1761. <https://doi.org/10.1212/WNL.44.9.1759>

- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA, USA: MIT Press.
- Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42(1–3), 1–22. [https://doi.org/10.1016/0010-0277\(92\)90038-J](https://doi.org/10.1016/0010-0277(92)90038-J)
- Lewandowsky, S., & Farrell, S. (2008). Phonological similarity in serial recall: Constraints on theories of memory. *Journal of Memory and Language*, 58(2), 429–448. <https://doi.org/10.1016/j.jml.2007.01.005>
- Lochy, A., Domahs, F., Bartha, L., & Delazer, M. (2004). Specific order impairment in arabic number writing: A case-study. *Cognitive Neuropsychology*, 21(5), 555–575. <https://doi.org/10.1080/02643290342000618>
- Lochy, A., Domahs, F., & Delazer, M. (2004). A Case-study of access deficit to stored multiplication facts: Discrepancy between explicit and implicit tasks. *Cortex*, 40(1), 153–154. [https://doi.org/10.1016/S0010-9452\(08\)70930-4](https://doi.org/10.1016/S0010-9452(08)70930-4)
- Mark-Zigdon, N., & Katzoff, A. (2015, February). *Best conditions for memory of multiplication facts*. Paper presented at the 2nd Conference on Cognition Research of the Israeli Society for Cognitive Psychology, Akko, Israel.
- Martin, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*, 33(1), 83–111. <https://doi.org/10.1006/jmla.1994.1005>
- McCrink, K., & Spelke, E. (2010). Core multiplication in childhood. *Cognition*, 116(2), 204–216. <https://doi.org/10.1016/j.cognition.2010.05.003>
- Monsell, S. (1987). On the relation between lexical input and output pathways for speech. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production* (pp. 273–311). London, United Kingdom: Academic Press.
- Nelson, D. L., Brooks, D. H., & Borden, R. C. (1974). Effects of formal similarity: Phonetic, graphic, or both? *Journal of Experimental Psychology*, 103(1), 91–96. <https://doi.org/10.1037/h0036821>
- Nickels, L. (1997). *Spoken word production and its breakdown in aphasia*. Hove, United Kingdom: Psychology Press.
- Nickels, L., Howard, D., & Best, W. (1997). Fractionating the articulatory loop: Dissociations and associations in phonological recoding in aphasia. *Brain and Language*, 56, 161–182. <https://doi.org/10.1006/brln.1997.1732>
- Noël, M. P., Désert, M., Aubrun, A., & Seron, X. (2001). Involvement of short-term memory in complex mental calculation. *Memory & Cognition*, 29(1), 34–42. <https://doi.org/10.3758/BF03195738>
- Oberauer, K. (2009). Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory & Cognition*, 37(3), 346–357. <https://doi.org/10.3758/MC.37.3.346>
- Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55(4), 601–626. <https://doi.org/10.1016/j.jml.2006.08.009>
- Oberauer, K., & Lange, E. B. (2008). Interference in verbal working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration. *Journal of Memory and Language*, 58(3), 730–745. <https://doi.org/10.1016/j.jml.2007.09.006>

- Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., . . . Ward, G. (2018). Benchmarks for models of short term and working memory. *Psychological Bulletin*, *144*(9), 885-958. <https://doi.org/10.1037/bul0000153>
- Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19*(5), 779-819. <https://doi.org/10.3758/s13423-012-0272-4>
- Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*(2), 227-252. <https://doi.org/10.1016/j.cognition.2009.09.007>
- Pajak, B., Creel, S. C., & Levy, R. (2016). Difficulty in learning similar-sounding words: A developmental stage or a general property of learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1377-1399. <https://doi.org/10.1037/xlm0000247>
- Patterson, K., & Shewell, C. (1987). Speak and spell: Dissociations and word-class effects. In M. Coltheart, G. Sartori, & R. Job (Eds.), *The cognitive neuropsychology of language* (pp. 273-294). Hove, United Kingdom: Erlbaum.
- Pesenti, M., Seron, X., & van der Linden, M. (1994). Selective impairment as evidence for mental organisation of arithmetical facts: BB, a case of preserved subtraction? *Cortex*, *30*(4), 661-671. [https://doi.org/10.1016/S0010-9452\(13\)80242-0](https://doi.org/10.1016/S0010-9452(13)80242-0)
- Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547-555. <https://doi.org/10.1016/j.neuron.2004.10.014>
- Posner, M. I., & Konick, A. F. (1966). On the role of interference in short-term retention. *Journal of Experimental Psychology*, *72*(2), 221-231. <https://doi.org/10.1037/h0023458>
- Runquist, W. N. (1970). Acoustic similarity among stimuli as a source of interference in paired-associate learning. *Journal of Experimental Psychology*, *83*(2, Pt.1), 319-322. <https://doi.org/10.1037/h0028547>
- Runquist, W. N. (1971). Stimulus coding and interference in paired-associate learning. *Journal of Experimental Psychology*, *87*(3), 373-377. <https://doi.org/10.1037/h0030535>
- Shalev, N., Ophir, E., Gvion, A., Gil, M., & Friedmann, N. (2014, February). *Dissociations between production processes of numbers and words*. Paper presented at the first Conference on Cognition Research of the Israeli Society for Cognitive Psychology, Akko, Israel.
- Shallice, T., Rumiati, R. I., & Zadini, A. (2000). The selective impairment of the phonological output buffer. *Cognitive Neuropsychology*, *17*(6), 517-546. <https://doi.org/10.1080/02643290050110638>
- Shallice, T., & Warrington, E. K. (1977). Auditory-verbal short-term memory impairment and conduction aphasia. *Brain and Language*, *4*, 479-491. [https://doi.org/10.1016/0093-934X\(77\)90040-2](https://doi.org/10.1016/0093-934X(77)90040-2)
- Simon, O., Mangin, J. F., Cohen, L., Le Bihan, D., & Dehaene, S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron*, *33*(3), 475-487. [https://doi.org/10.1016/S0896-6273\(02\)00575-5](https://doi.org/10.1016/S0896-6273(02)00575-5)
- Temple, C. M., & Sherwood, S. (2002). Representation and retrieval of arithmetical facts: Developmental difficulties. *The Quarterly Journal of Experimental Psychology, Section A*, *55*(3), 733-752. <https://doi.org/10.1080/02724980143000550>

- Vakil, E., & Blachstein, H. (1997). Rey AVLT: Developmental norms for adults and the sensitivity of different memory measures to age. *The Clinical Neuropsychologist*, *11*(4), 356-369. <https://doi.org/10.1080/13854049708400464>
- Vallar, G. (2006). Memory systems: The case of phonological short-term memory. A festschrift for Cognitive Neuropsychology. *Cognitive Neuropsychology*, *23*(1), 135-155. <https://doi.org/10.1080/02643290542000012>
- Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157-166. <https://doi.org/10.1016/j.jml.2006.03.007>
- van Harskamp, N. J., & Cipolotti, L. (2001). Selective impairments for addition, subtraction and multiplication: Implications for the organisation of arithmetical facts. *Cortex*, *37*(3), 363-388. [https://doi.org/10.1016/S0010-9452\(08\)70579-3](https://doi.org/10.1016/S0010-9452(08)70579-3)
- Wechsler, D. (1997). *WAIS-III: Wechsler Adult Intelligence Scale - Third Edition*. San Antonio, TX, USA: Psychological Corporation.
- Whalen, J., McCloskey, M., Lindemann, M., & Bouton, G. (2002). Representing arithmetic table facts in memory: Evidence from acquired impairments. *Cognitive Neuropsychology*, *19*(6), 505-522. <https://doi.org/10.1080/02643290244000086>
- Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, *55*(1), 235-269. <https://doi.org/10.1146/annurev.psych.55.090902.141555>
- Zbrodoff, N. J., & Logan, G. D. (2005). What everyone finds: The problem size effect. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 331–345). New York, NY, USA: Psychology Press.

## Appendix A: The Origin and the Measurement of Interference

In the present section we use DL's performance to evaluate 7 different cognitive models for the representation of arithmetic facts in memory. Each such cognitive model implies what makes two facts similar to each other and interfere with each other. Consequently, each model can predict how well DL should learn a given set of facts. Here, we evaluate each model's predictions versus DL's actual performance.

The cognitive assumptions of each model, and the corresponding method for computing similarity between two facts, are listed in [Table A.1](#) and depicted in [Figure A.1](#). Model 1 assumes that memorizing a multiplication fact involves activation of the fact's digits, and model 2 assumes activation of associations between digits. Models 3 and 4 are similar to 1 and 2, except that they assume that *number words* (rather than digits), or their associations, are activated, in line with the view that multiplication facts are stored verbally (Dehaene, 1992; Dehaene & Cohen, 1995). Model 5 adds the assumption that the representation distinguishes between the words of the operands and the words of the result. Model 6 assumes that only the operands count towards similarity. Last, model 7 focuses on the representation of numbers as magnitudes using the Approximate Number System (Dehaene, 1992; Dehaene & Cohen, 1995), and considers the problem size – the result of multiplying the 2 operands (similar results were obtained when defining problem size as the average operand size). The similarity between two facts was defined as the negated distance between problem sizes, transformed to log (because quantities are assumed to be represented internally on a compressed scale, Dehaene, 1997; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004).

The seven models were compared using three methods. The first method was based on the idea that a similarity index derived from the correct model should correlate with DL's pre-training scores (essentially, this is what we described above in the analysis of pre-training score in the Results Section). The second method applies the same idea to DL's scores after training. The third method, also based on DL's post-training scores, considers each of the 4 sets of the trained facts: a simi-

ilarity index derived from a correct model should show low similarity for sets that DL learned well, and high similarity for sets that she did not learn well.

Table A.1

*Different Cognitive Models for Interference, and the Definition of a Similarity Index Derived From Each Model*

	Cognitive assumptions: Memorizing a multiplication fact involves activation of...	Definition of interference: Interference stems from the activation of...	Similarity computation: The similarity between two facts is defined as...	Similarity of sample pairs		
				7×6=42 7×4=28	3×4=12 3×8=24	6×4=24 6×9=54
1	the fact's digits: 2 operands, 1-2 result digits	identical digits in the 2 facts	the number of digits shared by the two facts	3	3	2
2 <sup>a</sup>	associations between the fact's digits	identical digit associations in the 2 facts	the number of digit pairs shared by the two facts	3	3	1
3	the fact's number words: 2 operands, result words <sup>b</sup>	identical words in the 2 facts	the number of words shared by the two facts	1	2	2
4	associations between the fact's words <sup>b</sup>	identical word associations in the 2 facts	the number of word pairs shared by the two facts	0	1	1
5	the fact's number words, also bound with their role as operand or result	identical words in the same roles in the 2 facts	the number of operand words or result words shared by the two facts	1	1 <sup>c</sup>	2 <sup>c</sup>
6	the operands	an identical operand	0 if no shared operands 1 if a shared operand exists	1	1	1
7	the problem size (magnitude)	facts with numerically-close results	$-\log( R1-R2 )$ , where R1 and R2 denote the results of the two facts.	$-\log(14)$	$-\log(12)$	$-\log(30)$

<sup>a</sup>This is the similarity index used by De Visscher and Noël (2014b), and here for grouping the facts in our training program.

<sup>b</sup>Two additional models may assume that representing a fact involves activation of the fact's digits, where each digit is bound to its decimal role in the number (e.g.,  $7 \times 4 = 28$  involves "7 units", "4 units", "2 decades", and "8 units"). These models were not included here, because they are almost equivalent with the models that assume activation of number words: for the range 1-100 (excluding 11-19), a number words directly maps to a digit in a decimal position.

<sup>c</sup>The repetition of the digit "4" in two different roles (operand in one exercise, result in the other) does not count towards similarity.

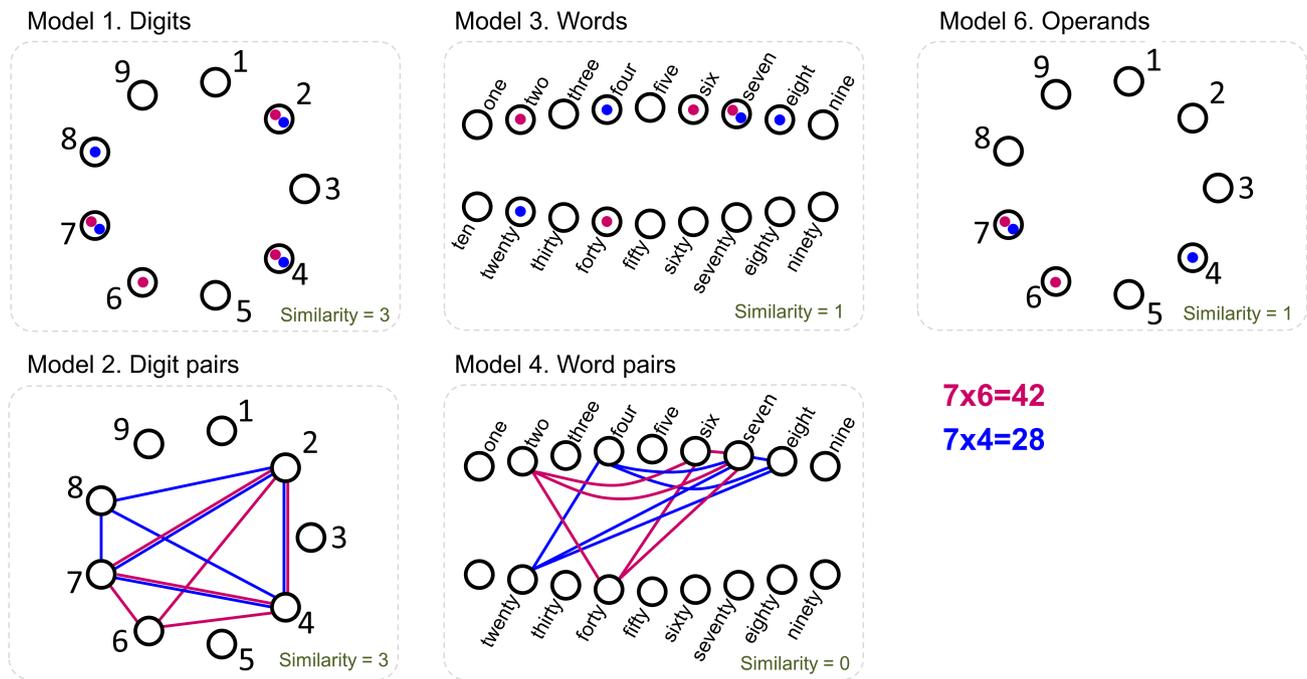


Figure A.1. Possible cognitive representations of multiplication facts, each implying a different source of interference and a different index of similarity between two facts. The models are explained in detail in the text. Black circles denote the representations of words or digits in long-term memory. The colored dots/lines denote the activation of these representations for two specific multiplication facts. In all cases, the similarity between two facts (green text) is computed as the number of elements (circles or lines) activated by both facts.

### A.1. Evaluation of the 7 Models Based on the Pre-Training Scores

The first method to evaluate the six models was based on DL’s pre-training scores. We reasoned that DL should show lower pre-training knowledge for multiplication facts that have higher similarity with the rest of the multiplication table, and that this correlation should be the highest when using a similarity index derived from the correct model. Thus, we computed how well the fact-table similarity defined by each model actually predicted DL’s pre-training scores. We used the bootstrapping method described in the analysis of pre-training score in the Results Section, which examines whether the fact-table similarities (*Sim*) of facts with correct pre-training responses are lower than the *Sim* values of facts with incorrect pre-training responses. The only difference was that now, the similarity between a pair of facts (and consequently the fact-table similarity, *Sim*) was defined in a different way for each model (Table A.1). In this analysis, *p* reflects how well each model predicts the observed data. Model 6 was excluded from the analysis, because it yields a constant fact-table similarity for all facts.

Table A.2 [a] shows that the *Sim* values extracted from DL’s pre-training performance was significantly higher than chance for models 1, 2, 3 and 4. The differences between the four models were fairly small, and in the absence of statistical comparison our data cannot arbitrate between them. The two remaining models failed to predict DL’s pre-training scores.

Table A.2

*Model Comparison.* To evaluate the six cognitive models for the representation of arithmetic facts, we compared DL's accuracy for each fact versus each model's definition of *Sim* – the similarity of that fact with other facts. Analysis (a) compared DL's pre-training accuracy against each fact's similarity with all other facts in the multiplication table. Analysis (b) compared DL's accuracy during training against each fact's similarity with the 3 other facts in the same training set. In both analyses, models 1–4 made better predictions than the other models

The model used to define <i>Sim</i>	Average fact-table similarity ( <i>Sim</i> ) for...		Sim(correct) < Sim(incorrect)	
	Correct responses	Incorrect responses	1-tail <i>p</i>	<i>z</i>
<b>[a] Sim(fact-table) vs. pre-training accuracy (all facts)</b>				
Model 1: No. of identical digits	26.0	37.6	.002	2.7
Model 2: No. of identical digit pairs (used in this study to group the facts)	7.5	14.7	.003	2.5
Model 3: No. of identical number words	13.9	20.4	.008	2.3
Model 4: No. of identical word pairs	1.1	4.0	.005	2.4
Model 5: No. of identical words-in-role	12.2	13.7	.17	1.0
Model 7: Problem size proximity	-67.5	-79.2	– <sup>a</sup>	
<b>[b] Sim(fact-set) vs. the during-training test accuracy (only for trained facts)</b>				
Model 1: No. of identical digits	4.3	5.5	.007	2.6
Model 2: No. of identical digit pairs (used in this study to group the facts)	1.5	3.3	.004	2.8
Model 3: No. of identical number words	2.1	3.0	.03	2.0
Model 4: No. of identical word pairs	0.23	0.77	.008	2.7
Model 5: No. of identical words-in-role	1.8	2.5	.06	1.7
Model 7: Problem size proximity	-7.5	-6.7	.10	1.3

<sup>a</sup>In this model, contrary to the similarity-induced-interference prediction, facts with correct responses were *more* similar to each other than the incorrect-response facts.

## A.2. Evaluation of the 7 Models Based on DL's Training Results

The second method to evaluate the models examined how well each model accounts for DL's **post-training** scores. For each model, we considered the model's definition of a fact's similarity with the 3 other facts in the same training set, and examined whether this similarity correlated with DL's accuracy for that fact in the weekly tests administered during the training period (similar results were obtained for the 2-month follow-up). The correlation was computed using the bootstrapping method described above. The results (Table A.2 [b]) were essentially the same as those obtained for the pre-training performance: models 1–4 significantly predicted DL's during-training accuracy, whereas models 5 and 7 did not.

A third method, which also used DL's post-training scores, was based on the idea that a similarity index derived from the correct model should show high similarity for the set of facts learned during the second week, in which DL performed poorly, and lower similarities for the other sets, in which DL performed better. For this analysis, the similarity between two facts was defined according to Table A.1, and the within-set similarity of a set of four facts was computed by summing the pairwise similarities of all six possible fact pairs in that set.

With this method, we could also examine the shared-operands model (model 6). Table A.3 indicates that this model did not agree with DL's performance: the within-set similarity of week 2, in which DL's performance was poor, was identical with the within-set similarity of week 4 (and nearly identical with week 3), in which she performed well. The words-in-role model (#5) too was not in good agreement with DL's performance. The remaining models (the digit-based, word-based, and quantity-based models) seem to be in reasonable agreement with DL's actual post-training performance. Nevertheless, the absence of a common similarity scale does not allow for a direct comparison between the models.

Table A.3

*Model Comparison*

DL's performance	Training week no.			
	1	2	3	4
	Good	Bad	Good	Good
<b>Facts trained during this week</b>	4×4=16	7×4=28	8×8=64	9×6=54
	8×3=24	7×6=42	9×7=63	6×5=30
	8×7=56	8×4=32	6×2=12	8×5=40
	5×3=15	9×4=36	8×6=48	7×5=35
<b>Model<sup>a</sup></b>				
Model 1: No. of identical digits	6	13	8	10
Model 2: No. of identical digit pairs <sup>b</sup>	0	9	3	4
Model 3: No. of identical number words	3	8	3	5
Model 4: No. of identical word pairs	0	2	0	1
Model 5: No. of identical words-in-role	2	6	3	5
Model 6: Common operand exists	2	4	3	4
Model 7: Problem size proximity	-131	-46	-171	-77

<sup>a</sup>The six cognitive models for the representation of arithmetic facts were evaluated by comparing each model's within-set similarity per training week versus DL's actual performance. Models 1, 2, 3, and 4 predict high interference in set #2 and low interference for the other sets, in agreement with DL's training results. <sup>b</sup>Used in this study to group the facts.

Altogether, these analyses suggest that models 1, 2, 3, and 4 – the models that assume that a fact is represented by activating single digits or number words, or pairs of digits or number words – make good predictions of DL's performance. The predictions of the other models were not as good, including the model that separates operand words from result words (5), the model that considers only the operand similarity (6), and the model that considers the similarity of problem size (7).

We acknowledge the limitations of these results: we did not offer a direct statistical comparison between the models, and the design of our intervention program did not consider all models equally (the facts were grouped into sets for DL's training based on similarity as defined by model 2). Future studies may elaborate more deeply on the precise representation of arithmetic facts.