12-6-2021

From counting to retrieving: Neural networks underlying alphabet arithmetic learning

Wim Fias
Universiteit Gent

Muhammet Ikbal Sahan
Universiteit Gent

Daniel Ansari
Western University, daniel.ansari@uwo.ca

Ian M. Lyons
Georgetown University

From Counting to Retrieving: Neural Networks Underlying Alphabet–Arithmetic Learning

Wim Fias[1*], Muhammet Ikbal Sahan[1*], Daniel Ansari[2], and Ian M. Lyons[3]

Abstract

■ This fMRI study aimed at unraveling the neural basis of learning alphabet–arithmetic facts, as a proxy of the transition from slow and effortful procedural counting-based processing to fast and effortless processing as it occurs in learning addition arithmetic facts. Neural changes were tracked while participants solved alphabet–arithmetic problems in a verification task (e.g., F + 4 = J). Problems were repeated across four learning blocks. Two neural networks with opposed learning-related changes were identified. Activity in a network consisting of basal ganglia and parieto-frontal areas decreased with learning, which is in line with a reduction of the involvement of procedure-based processing. Conversely, activity in a network involving the left angular gyrus and, to a lesser extent, the hippocampus gradually increases with learning, evidencing the gradual involvement of retrieval-based processing. Connectivity analyses gave insight in the functional relationship between the two networks. Despite the opposing learning-related trajectories, it was found that both networks become more integrated. Taking alphabet–arithmetic as a proxy for learning arithmetic, the present results have implications for current theories of learning arithmetic facts and can give direction to future developments. ■

INTRODUCTION

Being able to perform mental arithmetic is a cognitive skill of utmost importance in today's society. Although the basis for this skill may partly build on phylogenetic scaffolds of basic number representation (Nieder & Dehaene, 2009; Nieder, 2005), it needs substantial training through formal education and practice to reach the skill levels adapted to societal expectations, suggesting that the human brain is not built for learning arithmetic. In the absence of dedicated arithmetic learning mechanisms, arithmetic learning has to be achieved through the exploitation of general learning systems that subserve the acquisition of declarative knowledge and procedural skill. A detailed understanding of the involvement of these well-established declarative and procedural learning systems is much needed to make progress in understanding the neurocognitive basis of learning arithmetic as an important aspect and building block of mathematical cognition.

The ultimate goal of arithmetic training is to learn the basic arithmetic problems to such an extent that they can be effortlessly retrieved from memory. Such retrieval may arise through rote memorization or via elaborate practice with executing procedural strategies (or perhaps some combination thereof). For instance, before multiplication facts can be retrieved from memory, they are solved as repeated addition (Lemaire & Siegler, 1995). Or, at the beginning of learning, before children can easily produce 7 as the answer to 4 + 3, they apply finger counting procedures (4 + 3 = 4 + 1 + 1 + 1; Butterworth, 2005).

Whereas former research has laid out the general neural networks that play a role in arithmetic learning (Zamarian, Ischebeck, & Delazer, 2009) and the transition from effortful controlled processing to resource-minimal processing has begun to be addressed (Ischebeck, Zamarian, Egger, Schocke, & Delazer, 2007), additional work is necessary to substantiate and further specify the underlying dynamics of what happens to the procedural system and/or whether a separate retrieval-based system is created. Moreover, it remains to be determined how the transition is embedded at the neural level: Are new functional neural circuits developed, or are the original circuits reshaped to operate in a less effortful manner? Another outstanding question is how these circuits cooperate? Does one take over, or do they operate in parallel, and if so, do they mutually influence one another or do they run independently? The current study was designed to address these specific questions.

The most straightforward arithmetic operation to be studied if one wants to empirically address the above questions is addition, because it has two advantages. First, the procedural strategy that is used in initial stages of learning to add is well specified, in the sense that it is based on counting, often visible through overtly executed finger counting (Butterworth, 2005). Second, there is a clear behavioral signature of procedural learning in addition as it

[1]Ghent University, Belgium, [2]University of Western Ontario, Canada, [3]Georgetown University, Washington, D.C.
*Contributed equally.

 Journal of Cognitive Neuroscience X:Y, pp. 1–18
https://doi.org/10.1162/jocn_a_01789

unfolds. The addend size effect, which refers to the fact that RTs are longer with increasing addends, is large at initial stages of learning (around 400 msec per increment) and substantially reduces to a minimal size of a few tens of milliseconds per increment (Groen & Parkman, 1972). On the other hand, the study of addition is complicated by the fact that the transition from counting to retrieval occurs at ages too young for easily using neuroscientific techniques like fMRI.

Our strategy was to investigate the learning process of solving alphabet–arithmetic problems (Logan & Klapp, 1991) in adults. Alphabet–arithmetic problems are addition problems in which a number has to be added to a letter, with the number specifying the number of steps that have to be moved along the alphabetic sequence. As an example, K + 7 = R or P + 4 = T. Typically, even adult participants do not know the answer to alphabetic problems until they start explicitly learning it. Importantly, initial attempts to solve arithmetic problems necessarily rely on a well-specified counting-based procedural strategy, in which the number of steps along the alphabet are made as a function of the digit presented. As each additional counting step requires time to execute, performance is characterized by a strong addend size effect. Then, because executing the counting procedure is cognitively demanding, participants will soon rely on memory retrieval. This is empirically verifiable by the fact that the effect of addend size reduces substantially or completely disappears as learning progresses, thereby establishing the same behavioral signature as the one observed when children learn to solve addition problems. Moreover, with an adequate combination of number of problems and frequency of presentation, the transition from effortful counting to retrieval-based processing can be induced within the duration of a typical 1-hr fMRI session.

Counting can be expected to rely on procedure-based neural networks that comprise frontostriatal regions known to subserve skill acquisition (Poldrack & Gabrieli, 2001) and sequence learning (Gheysen, Van Opstal, Roggeman, Van Waelvelde, & Fias, 2011). The network supporting memory retrieval is expected to comprise the angular gyrus (AG), known to be involved in conceptual knowledge in general (Seghier, 2013), including arithmetic facts (Grabner et al., 2009; Ischebeck et al., 2007; Delazer et al., 2005). In addition, the medial temporal lobes may be involved, especially because they support initial stages of learning, although their role is still a matter of discussion. Some have found involvement of the hippocampus (HC; De Smedt, Holloway, & Ansari, 2011), whereas others did not (Ischebeck et al., 2007).

At the cognitive level, different predictions can be made, depending on the theoretical position. A first view, the two-systems view, builds on the idea that repeated procedural practice leads to the gradual creation of a memory trace that allows rapidly retrieving the solution to a problem (Compton & Logan, 1991; Logan, 1988). Hence, with learning, a qualitatively different system is engaged that exists independently from and in addition to the procedural system. An important argument in support of this view is the fact that the addend size effect as observed in addition substantially decreases or even disappears with learning. The reduction of the addend size effect is considered so big that it is taken as an indication that the time-consuming counting strategy has been replaced by an effortless retrieval strategy whereby the answer to a problem is directly retrieved from memory.

A second view, on the other hand, proposes that training leads to optimization of the underlying procedures, such that they become more time-efficient and consume less resources (Uittenhove, Thevenot, & Barrouillet, 2016; Barrouillet & Thevenot, 2013). For instance, finger or verbal counting procedures are thought to be executed without actual muscular activity, but rather to occur completely internally, and hence quicker. This one-system view relies on the fact that the addend size effect becomes smaller with learning but does not completely disappear.

At a neural level, the crucial issue is whether learning is accompanied by the engagement of new functional neural circuits that had not previously been engaged but now do become involved in solving mental arithmetic problems. Previous work suggests that this is the case. Ischebeck et al. (2007) investigated the neural consequences of repetitively solving complex multidigit multiplication problems. They found that increasing the number of repetitions of a problem increased activation in the left AG (L AG), and decreased activation in parieto-frontal regions as well as the caudate nucleus. This provides evidence in support of the two-systems view, as it establishes opposite learning-dependent changes. In an fMRI study in children (aged 10–12 years), hippocampal activity was observed by De Smedt et al. (2011) in addition and subtraction problems, which, based on their small problem size, were expected to be solved by retrieval.

Whereas the above studies support the two-systems view, a study by Cho, Ryali, Geary, and Menon (2011) provides support for the one-system view. They investigated children aged 7–9 years. At this age, learning to add is still at the beginning. Some children may already have learned to retrieve addition facts whereas others are still in the counting stage. When comparing the brain activations of counters and retrievers in a univariate way, no differences were found (apart from activation in the left ventrolateral pFC in retrievers), suggesting that the same circuits underlie addition irrespective of counting or retrieving. However, substantial differences between counters and retrievers were established in the multivariate patterns in the HC and the parietal cortex. The fact that counters and retrievers do not differ in terms of which parts of the brain they use, but rather in how these regions are used (as reflected in the multivariate patterns), suggests that learning leads to a reorganization of the neural system that learners start from, rather than the recruitment of previously unused neural circuits.

In the current study, we tracked the neural changes that accompany the learning and acquisition of alphabet–arithmetic problems. Based on earlier behavioral studies, it is known that an initial time-consuming counting-based solution process is replaced by efficient effortless processing. However, it is not yet known whether this reflects learning-induced optimization of the counting procedures or, alternatively, a transition toward direct memory retrieval that bypasses counting procedures altogether. For this reason, we presented participants with alphabet–arithmetic problems and tracked both behavioral performance and neural activity across four consecutive blocks in which the same arithmetic problems were repeated.

To evaluate the contribution of procedure- and retrieval-based problem-solving, we examined problems with different addend sizes (4 and 7; e.g., J + 4 or J + 7). Based on previous behavioral work using alphabet–arithmetic (Logan & Klapp, 1991), we expected participants to initially rely on the alphabetical version of a counting strategy, that is, by rehearsing the alphabetic sequence starting from the initial letter stimulus (e.g., J + 4 would be solved by rehearsing "J, K, L, M, N"). This strategy yields slower response times for larger addends because one must rehearse a greater number of letters. As previous work has demonstrated, we would expect, at least initially, J + 4 to produce substantially faster response times than J + 7. With repeated practice, however, this addend effect tends to decrease or even vanish altogether. The crucial question we address here is whether the reduction in the addend effect occurs because of direct, rote retrieval ("J + 4 = N"), in line with the two-systems view, or to increased efficiency of the rehearsal process itself (such that the cost of another letter to the rehearsal chain becomes vanishingly small), in line with the one-system view. Prior behavioral work has struggled to distinguish these possibilities. Here, we brought to bear multiple analysis techniques using fMRI data to identify neural networks supporting procedure-based and retrieval-based solution strategies, based on the learning-related evolution of activity. More specifically, procedure-based neural activity should decrease with learning as the application of the counting procedures become increasingly efficient. Conversely, retrieval-based networks should follow an opposite temporal pattern: As retrieval is impossible initially and can only emerge with learning, activity in retrieval-related brain areas should increase with learning. Note that these patterns are not mutually exclusive: Some brain areas might show a "procedural" pattern; others might show a "retrieval" pattern; if so, this might indicate both types of learning are at play.

In addition to traditional univariate analyses, we assessed the connectivity among brain regions in procedural and retrieval networks, allowing us to identify learning-related changes both between and within these two types of learning networks. In this way, we sought to examine changes in the internal coherence of procedural and retrieval networks, and to evaluate whether these networks become increasingly integrated or dissociated as a function of learning. With respect to within-network connectivity, we expected the strength of the correlations between regions within the network would covary with how strongly the network contributes to learning. Hence, we expected connectivity within the procedural network to be strong at the beginning of learning and to then decrease, whereas we expected the connectivity between regions in the retrieval network (if such a network emerges) to show the opposite pattern. With respect to the between-networks connectivity, increased connectivity between networks with learning would indicate greater integration, whereas decreased between-networks connectivity would indicate greater dissociation.

METHODS

Participants

Twenty-eight healthy right-handed volunteers participated after having given written informed consent. Three participants were discarded from analyses because of excessive head motion (i.e., exceeding 1-mm translation and 1deg rotational on consecutive TRs). The study was approved by the Health Science Research Ethics Board at the University of Western Ontario.

Stimuli and Materials

Participants were randomly assigned one out of four learning sets. Table 1 presents a complete list of alphabet–arithmetic problems that were used per learning set.

An alphabet–arithmetic problem consisted of a letter operand, the operator sign +, a digit addend, the equal sign "=" followed by a letter as proposed solution (e.g., "F + 4 = J"). Participants' task was to verify whether the operand letter plus the number of steps in the alphabet indicated by the number addend equals the proposed

Table 1. Alphabet–Arithmetic Problems Per Learning Set

	Set 1		*Set 2*		*Set 3*		*Set 4*	
	Letter	*Digit*	*Letter*	*Digit*	*Letter*	*Digit*	*Letter*	*Digit*
HF	U	+0	U	+0	U	+0	U	+0
	F	+4	I	+4	F	+4	H	+4
	G	+7	J	+7	G	+7	I	+7
LF	V	+0	V	+0	V	+0	V	+0
	I	+4	F	+4	H	+4	F	+4
	J	+7	G	+7	I	+7	G	+7
	K	+4	K	+4	J	+7	J	+7
	L	+7	L	+7	K	+4	K	+4

letter. For instance, for the problem "F + 4 = J," participants had to indicate whether or not J is four steps further down the alphabet from F.

In every set, letter operands were F, G, J, K, I, U, and V and depending on the set L (Sets 1 and 2) or H (Sets 3 and 4). Digit addends were 0, 4, and 7. Problems with digit addend 0 constituted the control condition, because no procedure or retrieval has to be applied. The experimental condition consisted of problems with digit addends 4 or 7, with the addend size manipulated to introduce a difference in the number of steps to be made in the alphabet. The proposed solution was either the correct letter or the letter following the correct letter in the alphabet.

The frequency of occurrence of the problems was manipulated, with problems in the high-frequency condition being presented twice as often as the problems in the low-frequency condition. The frequency manipulation intended to spread the learning trajectories of the individual problems across time with the ultimate goal of aligning neural changes to the point of transition from procedure to retrieval (see Ison, Quian Quiroga, & Fried, 2015, for a similar approach). Unfortunately, this did not work because the transition points turned out not to be clearly demarcated in time. This may be related to the fact that more repetitions are needed before stable transition to retrieval is made. As proposed by Rickard (1997), at least 60 trials are needed to achieve automatic retrieval. Therefore, the frequency manipulation was not taken into account in the analyses.

The alphabet–arithmetic problems were organized in four learning blocks consisting of three subblocks each. Each subblock consisted of 22 trials in which each of the high-frequency problems was presented 4 times (2 times with the correct solution and 2 times with the incorrect solution) and each of the low-frequency problems 2 times (once with the correct solution and once with the incorrect solution).

The four learning blocks were followed by a postlearning phase consisting of a subblock of 22 trials, which was organized in the same way as the learning subblocks but with novel letters: M, N, O, P, Q, R, D, E. There were two control problems (of type: D + 0, E + 0), three learning problems with addend 4 (of type: M + 4, O + 4, R + 4), and three learning problems with addend 7 (of type: N + 7, P + 7, Q + 7). One problem of each of the three addend distances (i.e., D + 0, M + 4, N + 7) occurred 4 times in the high-frequency condition whereas the remaining five problems occurred 2 times in the low-frequency condition. As with the learning blocks, the novel trials were also presented along with the true and false solutions that the participants were to verify. This postlearning block was added to distinguish slow effects related to the participant (e.g., fatigue, boredom,…) or scanner (signal drift) from true learning effects.

The stimuli were presented inside the scanner via a Brainlogics 200 MR digital projector visible via a mirror attached to the head coil, with a viewing distance of 120 cm. Each problem was presented as a capital letter, a plus symbol (+), a numeral (digit addend), an equals sign (=), and another capital letter (true/false solution), in Calibri 36 pt. font. The components of the problem were separated by single spaces. Each time a learning block was completed, participants were asked to report whether they retrieved the solutions to the six problems (thus, not the control problems U + 0 and V + 0) from memory. The strategy report questions were presented in a centrally aligned fashion on both the horizontal and vertical axes. The questions were stated as follows: "In the most recent block of trials, did you recall this problem from memory?" along with which the problem was presented. These problems were presented in the same format as in the learning blocks, but without the solution (e.g., "L + 4 =") upon which the participants were to report their strategy.

Experimental Procedure

Participants completed four learning blocks (Blocks 1–4) and one postlearning block (Block 5). There were 66 trials inside a learning block and 22 trials inside the postlearning block. All stimuli were presented at the center of the screen in white, unless specified otherwise, against a black background. Each trial started with a fixation square (■) presented for 600 msec. The alphabet–arithmetic problem was then presented with a response deadline of 10 sec. Participants were instructed to determine whether the equation was either true or false by pressing the left or right button pressed with the index and middle finger, respectively, of the right hand. As a way of providing feedback, for 1 sec, the equation turned into either blue or red when responded correctly or incorrectly, respectively. An intertrial interval of 1 sec was administered during which the fixation square was presented.

Upon completion the 66 trials of a learning block, a fixation interval of 15 sec was presented during which the fixation square appeared. This initiated the phase in which strategy report questions were asked about the recently finished learning block. There were six questions, presented one at a time, that were addressing whether participants solved the six learning problems from memory. First, instructions were displayed for 2 sec indicating the start of the strategy report phase. Next, a fixation square was presented for 600 msec that was then followed by the strategy report question. No response deadline was administered. Participants were to indicate whether the presented problem was retrieved from memory or not by pressing the left or right button, respectively. The interval between strategy questions was 1100 msec. Participants answered six strategy report questions, one for each learning problem. After another fixation interval of 14 sec, participants were given a break before the next block started. The next block was initiated by the experimenter after the participants indicated they were ready to proceed. Stimulus presentation and response collection were controlled using E-Prime (Psychology Software Tools; Figure 1).

Scanning Procedure

Scanning was performed in a 3T Siemens Magnetom Prisma, using a 32-channel head coil. First, a 3-D high-resolution T1-anatomical image of the whole brain was acquired with 3-D magnetization prepared rapid gradient echo (TR = 2500 msec, TE = 30, TI = 900 msec, 176 sagittal slices, voxel size = 1 × 1 × 1 mm (no skip), in-plane acquisition matrix = 256 × 256, field of view = 256 × 256 × 176 mm, flip angle = 9°). Next, whole-brain functional images were acquired using a T2*-weighted sequence sensitive to BOLD contrast (EPI: TR = 2600 msec, TE = 59 msec, flip angle = 80°, 44 axial slices, in-plane matrix = 64 × 64 [in-plane resolution = 3.3 × 3.3 mm], slice thickness = 3.3 mm with 0-mm skip; hence, field of view = 211.2 × 211.2 × 145.2 mm. Functional images were subsequently resampled to isotropic voxel size of 3 × 3 × 3 mm). Given the self-paced nature of the task, a variable number of functional images (*M* = 1033 scans, *SD* = 71) was acquired in a single run for each subject. The total scan duration was approximately 45 min. Padding was used around the head to reduce head motion.

Analysis

fMRI Preprocessing

The fMRI data were analyzed with statistical parametric mapping (SPM12; www.fil.ion.ucl.ac.uk/spm). The first five volumes of each EPI series allowed for magnetic saturation and were removed from the analysis. All functional images were spatially realigned to the mean image and were then temporally realigned to the middle slice. Next, the segment toolbox was used to segment structural images into gray matter, white matter, cerebrospinal fluid, bone, fat, and air by registering tissue types to tissue probability template maps (Ashburner & Friston, 2005). Bias corrected skull-stripped anatomical images were then generated to which the functional images were coregistered. These functional images were normalized to a standard EPI template in Montreal Neurological Institute (MNI) stereotaxic space and resampled at an isotropic voxel size of 3 mm by applying the forward deformation parameters that were obtained from the segmentation procedure. The normalized images were smoothed with a spatial filter of 8 mm FWHM. Finally, the high-resolution anatomical scan of each participant was spatially normalized using the same parameters that were applied for the normalization of the functional images.

Whole-Brain Analysis

We performed a general linear model at the first level for each participant on a trial-by-trial basis with separate events for condition and learning blocks leading to a model with two orthogonal dimensions: learning condition (learn [addends 4 or 7] or control [addend 0]) and learning block (1 through 5), creating a total of 10 conditions of interest (Learn Blocks 1–5 and Control Blocks 1–5). To model the hemodynamic response for each event, a stick function time-locked to its occurrence was convolved with a canonical hemodynamic response function

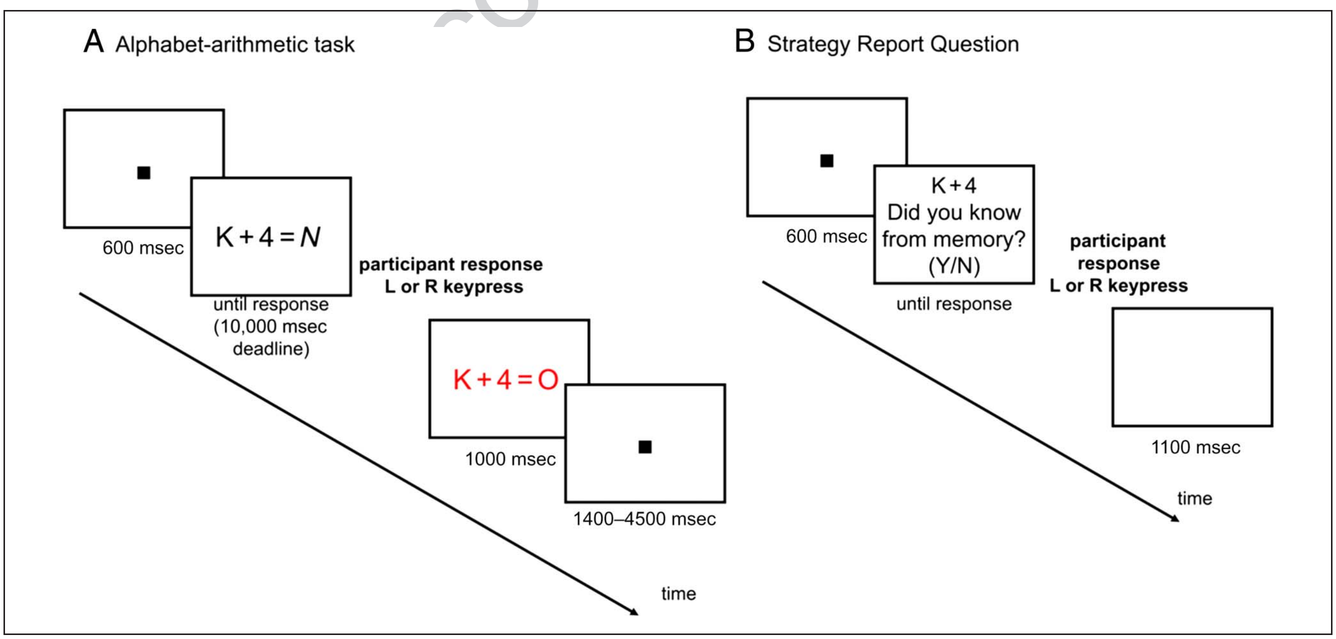

Figure 1. Task procedures. (A) Alphabet–arithmetic task. Participants performed four learning and one novel block of alphabet–arithmetic trials. (B) Strategy report questions. After each learning block, participants were instructed to report whether they solved the alphabet–arithmetic problems by retrieving the problems from their memory or not.

to form covariates in a general linear model (Friston, Glaser, Mechelli, Turner, & Price, 2003). In addition, the strategy report questions, fixations, and trials on which participants responded incorrectly and movement-related effects (three translational and three rotational) were included as covariates of no interest. High-pass filtering at a cutoff of 128 Hz and AR(1) serial correlation correction were also included in the analysis. The parameter estimates for each of the 10 conditions of the first level were entered into the second level by using a 2 × 5 factorial design (learn vs. control × Blocks 1 through 5) with participants as a random variable (Friston, Holmes, Price, Büchel, & Worsley, 1999). To achieve a corrected extent threshold of $p < .05$ at the cluster level (voxel level $p < .001$, uncorrected), a minimum cluster size of 30 voxels was used based on Monte Carlo simulations (afni.nimh.nih.gov/pub/dist/doc/program_help/3dClustSim.html), following Forman et al.'s study (1995).

Specific contrasts were then used to identify the procedure- and retrieval-based activations. Because procedure-based processing is maximal the first times that an alphabet–arithmetic problem has to be solved, it can be expected that procedure-related neural activity is stronger in Block 1 and Block 5 of the learning condition (as in these blocks' new problems have to be learned) compared to Block 4 (in which the problems have already been solved multiple times). Hence, the contrast $(\text{Block1}_{\text{Learn}} + \text{Block5}_{\text{Learn}}) - 2 \times \text{Block4}_{\text{Learn}}$ contains procedure-related neural activity. To retain those voxels that show this pattern of activity specifically in the learning condition, this contrast was exclusively masked with the same contrast in the control condition $[(\text{Block1}_{\text{Control}} + \text{Block5}_{\text{Control}}) - 2 \times \text{Block4}_{\text{Control}}]$ at a threshold of $p < .05$ uncorrected. For retrieval-based processing, an analogue approach but opposite reasoning was followed. As retrieval is expected to be strongest when the problems have been solved most frequently (i.e., in Block 4) compared to the first encounters with problems (i.e., in Block 1 and Block 5), we computed the contrast in the learning condition $[2 \times \text{Block4}_{\text{Learn}} - (\text{Block1}_{\text{Learn}} + \text{Block5}_{\text{Learn}})]$ and exclusively masked it with the same contrast in the control condition $[2 \times \text{Block4}_{\text{Control}} - (\text{Block1}_{\text{Control}} + \text{Block5}_{\text{Control}})]$ at a threshold of $p < .05$ uncorrected.

Effective Connectivity

The goal of this analysis was to examine the connectivity changes within and between the neural networks implicated in the alphabet–arithmetic task over the course of learning (from Block 1 to 4 and 5). To this end and using the initial parts of the psychophysiological interaction module in SPM12, we extracted time series, defined as the principal eigenvariate, for each subject's procedural and retrieval network regions. Unlike the mean, the principal eigenvariate does not assume homogenous response within a region and is therefore a more representative regional response (Friston, Rotshtein, Geng, Sterzer, & Henson, 2006). These time series were first deconvolved with the canonical hemodynamic response function in order to obtain the underlying neural activity (i.e., physiological variable). Note that this deconvolution step is necessary given that the interactions in the brain take place at a neural level and not at the hemodynamic level (Gitelman, Penny, Ashburner, & Friston, 2003). The neural activity was then multiplied with the task design vector contrasting the learning and control condition for each block separately (i.e., psychological variable). As a result of these steps, processed time series associated to learning were obtained for each subject (25), each ROI (16), and each Block (3: Block 1, Block 4, Block 5) separately.

For each subject, the Pearson correlation coefficients (r value) between each pair of the ROI time series were then calculated resulting in a 16 × 16 diagonally symmetrical pairwise connectivity matrix for each subject and each block. These r values were then grouped as to their membership to a network and averaged accordingly. More precisely, this grouping resulted in an average measure of connectivity for the *procedure-to-procedure* (P-to-P) and *retrieval-to-retrieval* (R-to-R) regions as within-network and *procedure-to-retrieval* (P-to-R) regions as a between-network. Note that P-to-R is the same as its reverse R-to-P, which corresponds to the opposite side of the diagonal (see Figure 7). These measures were then subjected to a repeated-measures ANOVA with network (P-to-P, R-to-R, P-to-R) and block (1, 4, and 5) as within-subject factors. All tests were conducted on Fisher z-transformed r values across participants [$z = \text{atanh}(r)$], and post hoc tests comparing the differences between conditions were done using paired t tests.

RESULTS

Behavioral Results

General Learning

In a first analysis, we investigated the overall effect of learning alphabet–arithmetic operations over the course of the task (see Figure 2). For this purpose, we compared the differences between learning (collapsing addend Distances 4 and 7) and control trials (Addend Distance 0) over the course of the task (Learning Blocks 1 through 4). Average RTs and error rates were each subjected to a repeated-measures ANOVA with learning blocks (1–4) and learning condition (learning vs. control trials) as within-subject factors.

The analysis revealed a significant main effect of learning condition, $F(1, 24) = 112.21$, $p < .001$, $\eta_p^2 = .82$, with longer latencies in the learning trials compared to the control trials. There was also a significant effect of learning block, $F(3, 22) = 81.02$, $p < .001$, $\eta_p^2 = .92$, with RTs decreasing from Block 1 to Block 4. The interaction term revealed that this RT decrease over blocks was

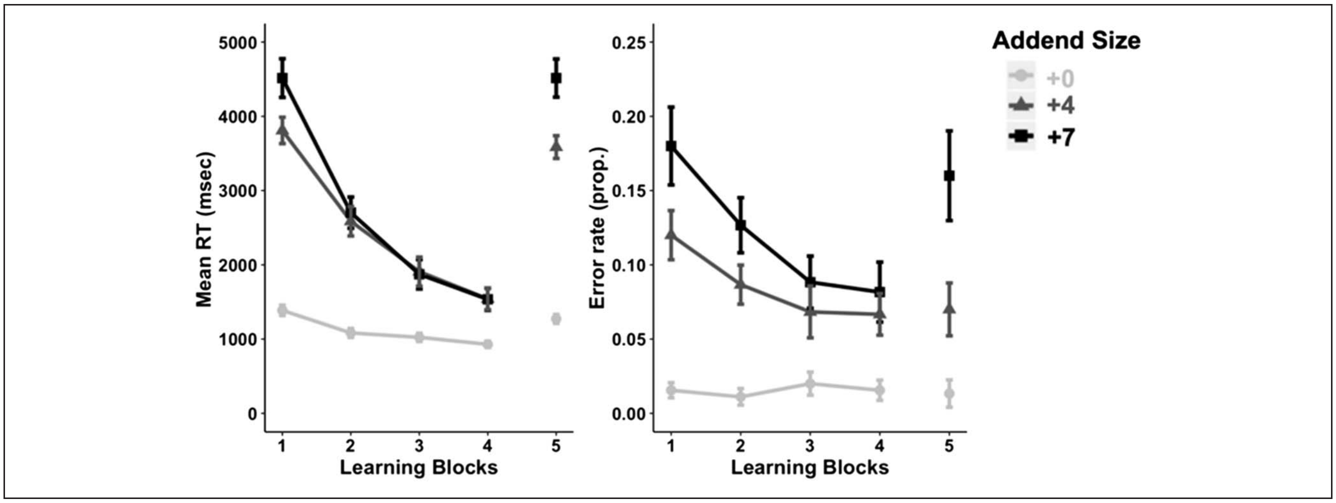

Figure 2. Performance by learning blocks and addend size. (A) RTs. (B) Error rates. Note that the lines are only connected for Blocks 1–4 (i.e., learning blocks). The disconnection with Block 5 is to indicate that this is a block with novel alphabet–arithmetic problems. Error bars represent the standard error of the mean.

significantly different between the learning conditions, $F(3, 22) = 59.95, p < .001, \eta_p^2 = .88$. In order to formalize this RT decrease over learning blocks, we fit linear regressions for each condition and compared the slopes between each condition. The RT decreases in the learning trials (slope = −857.20 msec, $SE = 58.19$; $t(24) = -14.73, p < .001$, 95% CI [−977.2929, −737.12) and control trials were both significant (slope = −143.18 msec, $SE = 18.99$; $t(24) = -7.54, p < .001$, 95% CI [−182.38, −103.98]). However, the decrease was significantly steeper in the learning trials ($M = -714$, $SE = 58$, $t(24) = -12.32, p < .001$, 95% CI [−833.63, −594.42]).

There were more errors in the learning trials compared to the control trials, $F(1, 24) = 46.40, p < .001, \eta_p^2 = .66$. Furthermore, the analysis revealed that the error rates also decreased by blocks, $F(3, 22) = 6.26, p < .005, \eta_p^2 = .46$. However, the decrease was different between conditions, as indicated by the interaction term, $F(3, 22) = 6.13, p < .005, \eta_p^2 = .46$. The error rate only decreased significantly in the learning trials (slope = −0.03, $SE = .005$; $t(24) = -5.17, p < .001$, 95% CI [−0.04, −0.01]) whereas the error rate remained flat in the control trials (slope = .0008, $SE = .003$; $t(24) = 0.23, p = .77$, 95% CI [−0.005, −0.007]).

Together, both RTs and error rates show that substantial learning occurred from Block 1 to Block 4.

Postlearning Block

By adding the postlearning block in which novel alphabet–arithmetic problems had to be learned, we wanted to force participants to restart the learning process, with the ultimate aim of distinguishing slow participant- or scanner-related effects from true learning effects.

Extending the previous analysis with the postlearning block as fifth block of the learning block variable in the above analysis revealed a significant main effect of learning condition, $F(1, 24) = 171.05, p < .001, \eta_p^2 = .88$; learning block, $F(4, 21) = 95.39, p < 0.001, \eta_p^2 = .95$; and a significant Learning Condition × Learning Block interaction, $F(4, 21) = 83.09, p < .001, \eta_p^2 = .94$. Planned comparisons were performed to test performance in the novel block in comparison to the first and fourth blocks, for each learning condition separately. The RTs in the novel block were significantly higher than the RTs in the fourth block both in learning ($M_{dif} = 2500.30$, $SE_{dif} = 135.2$, $t(24) = 18.49, p < .001$, 95% CI [2221.20, 2779.41]) and control conditions ($M_{dif} = 342.70$, $SE_{dif} = 47.65$, $t(24) = 7.19, p < .001$, 95% CI [244.35, 441.04]).

As for the error rates, the analysis revealed a significant main effect of learning condition, $F(1, 24) = 44.97, p < .001, \eta_p^2 = .65$; learning block, $F(4, 21) = 4.53, p < .01, \eta_p^2 = .46$; and a significant Learning Condition × Learning Block interaction, $F(4, 21) = 4.40, p < .05, \eta_p^2 = .46$. None of the paired comparisons between the novel block and Block 4 survived significance testing ($-2 < ts < 2$); hence, although there was a numerical indication of increased error rates in the learning condition from Block 4 to Block 5, it did not reach significance.

Addend Size

To evaluate whether the transition was made from procedural counting to retrieval to solve the alphabet–arithmetic problems, we investigated the effect of addend size over learning. We predicted that the addend size would have a pronounced effect at the beginning of learning and that the effect of addend size would gradually diminish and eventually disappear as learning proceeded. For this purpose, a 4 (learning block: 1–4) × 2 (addend size: 4 or 7) repeated-measures ANOVA was conducted. In line with the general learning effects, RTs significantly decreased over the course of the blocks, F(3,

22) = 70.77, $p < .001$, $\eta_p^2 = .91$. Although the main effect of addend size was only marginally significant, $F(1, 24) = 3.99$, $p < .06$, $\eta_p^2 = .14$, with numerically larger RTs for addend size 7 compared to 4, the effect of addend size was modulated over the course of the blocks, $F(3, 22) = 6.30$, $p < .005$, $\eta_p^2 = .46$. Planned comparisons revealed that the addend size effect was only significant in the first block ($M_{dif} = 705.34$, $SE_{dif} = 181.16$, $t(24) = 3.89$, $p < .005$, 95% CI [331.45, 1079.24]) whereas this difference disappeared from the second block onward ($-1 < t\text{s} < 1$). Nevertheless, the effect of addend size was reestablished in Block 5 with larger RTs for the large addend size compared to the small addend size ($M_{dif} = 929.66$, $SE_{dif} = 237.04$, $t(24) = 3.92$, $p < .005$, 95% CI [440.42, 1418.89]). These results clearly show that the effect of addend size substantially decreases from Block 1 to 4 and then increases again in Block 5.

As for the error rates, only the main effects of block, with decreasing errors over blocks, $F(3, 22) = 8.68$, $p < .005$, $\eta_p^2 = .54$, and addend size with more errors for the large compared to the small addend sizes, $F(1, 24) = 10.02$, $p < .005$, $\eta_p^2 = .30$, were significant. The Block × Addend Size interaction did not reach significance ($F < 2$). Extending this analysis with the novel block did not change the results.

Strategy Report

Figure 3 shows the number of problems that the participants report to have solved from memory, plotted by learning blocks. Reports of remembering significantly increased across blocks, $F(3, 22) = 20.11$, $p < .001$, $\eta_p^2 = .73$.

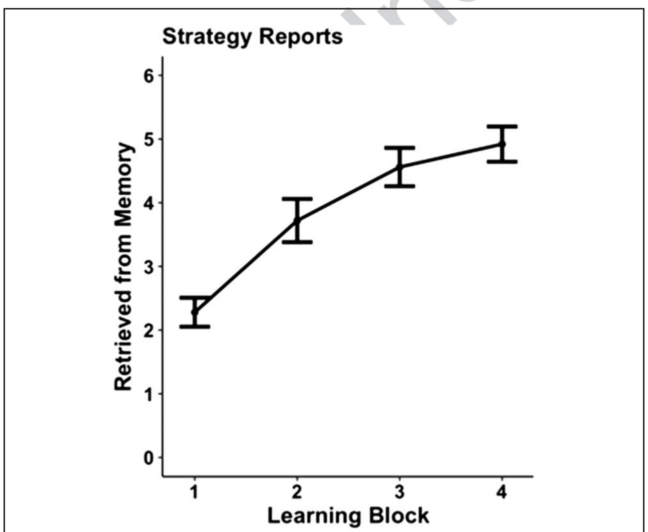

Figure 3. Strategy reports as a function of the learning blocks. Note that no strategy report questions were asked after Block 5. Error bars represent the standard error of the mean.

fMRI

Whole Brain

The statistical analyses were based on the behavioral learning patterns observed in the scanner. The general learning effect we observed, with faster and more correct responses across the four learning blocks compared to the control blocks, showed that the alphabet–arithmetic problems were retrieved from memory as learning proceeded. In addition, performance returned to its initial stage upon presentation of novel problems in the fifth block (see Figure 2). We hypothesized that regions involved in the learning process would show a similar pattern of a decrease over the course of the experiment with a rebound at the novel block. In other words, whereas Blocks 1 and 5 contain unlearned problems and therefore require procedure-based problem-solving (i.e., counting), Block 4 largely contains learned problems and therefore should require retrieval-based problem-solving. In order to reveal the regions involved in the learning process, we contrasted the blocks where performance was marked with counting (Blocks 1 and 5) with the block where performance was marked with retrieval (Block 4). To investigate the specificity of the procedure-based activations, we exclusively masked the procedure contrast in the learning condition $[(\text{Block1}_{\text{Learn}} + \text{Block5}_{\text{Learn}}) - 2 \times \text{Block4}_{\text{Learn}}]$ with the same contrast in the control condition $[(\text{Block1}_{\text{Control}} + \text{Block5}_{\text{Control}}) - 2 \times \text{Block4}_{\text{Control}}]$ at a threshold of $p < .05$ uncorrected. This contrast revealed a network comprising the bilateral basal ganglia and a widespread frontoparietal network including the bilateral frontal cortices and right superior parietal cortex (see Table 2A). The activity profiles in Figure 4A show that activity is strongest when problems begin to be learned (Blocks 1 and 5) and then decreases as learning progresses. In the control condition, no such pattern is present.

Conversely, we hypothesized that the reversed pattern would show the regions involved in the retrieving part of learning process. Specifically, we contrasted the block where problems were learned and therefore would involve retrieval-based processing (Block 4) with the blocks where problems were not learned yet and therefore required counting (Blocks 1 and 5). To investigate the specificity of the retrieval-based activations, we exclusively masked the retrieval contrast in the learning condition $[2 \times \text{Block4}_{\text{Learn}} - (\text{Block1}_{\text{Learn}} + \text{Block5}_{\text{Learn}})]$ with the same contrast in the control condition $[2 \times \text{Block4}_{\text{Control}} - (\text{Block1}_{\text{Control}} + \text{Block5}_{\text{Control}})]$ at a threshold of $p < .05$ uncorrected. This contrast identifies areas where activity is strongest after problems have been learned (Block 4) compared to when learning starts (Blocks 1 and 5) without learning related activity in the control condition. This contrast revealed a network of regions more posterior including the L AG. The results of the whole-brain contrasts are listed in Table 2B and depicted in Figure 4B.

Table 2. Procedural and Retrieval Networks Involved in the Alphabet–Arithmetic Task

Anatomical Region	*Hemisphere*	*Cluster Size*	*Z Score*	*MNI Coordinates*		
				x	*y*	*z*
(A) Procedure Network						
Frontal Lobe						
Precentral Gyrus	L	37	4.38	−54	5	20
Middle Frontal Gyrus	L	63	4.06	−24	2	56
	R	56	4.30	24	−1	53
			4.11	24	8	53
SMA	L	32	4.03	−3	5	68
Basal Ganglia						
Putamen	L	150	7.60	−18	11	2
			6.94	−17	11	8
			6.04	−24	11	−7
Caudate	R	88	5.90	18	17	8
			5.63	18	5	−1
			5.07	18	17	−4
Superior Parietal Lobe	R	53	4.13	18	−67	56
Middle Occipital Gyrus	L	33	4.39	−36	−82	32
(B) Retrieval Network						
Frontal Lobe						
Insula	R	378	4.80	39	8	11
			4.65	39	11	−7
			4.61	39	28	23
Insula	L	65	4.37	−39	−10	14
			3.54	−36	−5	14
			3.26	−39	−16	5

Table 2. (*continued*)

				MNI Coordinates		
Anatomical Region	*Hemisphere*	*Cluster Size*	*Z Score*	*x*	*y*	*z*
Anterior Insula	L	93	4.56	−33	14	−16
Median Cingulate Gyrus	L	40	4.01	−6	13	32
Parietal Lobe						
Precuneus	L	68	4.78	−15	−52	32
AG	L	35	3.81	−57	−58	38
Middle Temporal Gyrus	R	50	5.06	42	−61	5
Lingual Gyrus	R	337	5.64	24	−76	−7
			4.86	27	−67	−10
			4.69	30	−55	10

(A) Procedural network. List of areas that showed a significant activation related to counting with the learning condition [($Block1_{Learn}$ + $Block5_{Learn}$) − 2 × $Block4_{Learn}$] exclusively masked with the same contrast in the control condition [($Block1_{Control}$ + $Block5_{Control}$) − 2 × $Block4_{Control}$]. (B) Retrieval network. List of areas that showed a significant activation related to retrieval with the learning condition [2 × $Block4_{Learn}$ − ($Block1_{Learn}$ + $Block5_{Learn}$)] exclusively masked with the same contrast in the control condition [2 × $Block4_{Control}$ − ($Block1_{Control}$ + $Block5_{Control}$)]. Threshold level was set at $p < .001$ for the main contrast and at $p < .05$ for the exclusive mask. The anatomical labels for the MNI coordinates were obtained from the Anatomical Automatic Atlas Labeling toolbox for SPM12 (Tzourio-Mazoyer et al., 2002).

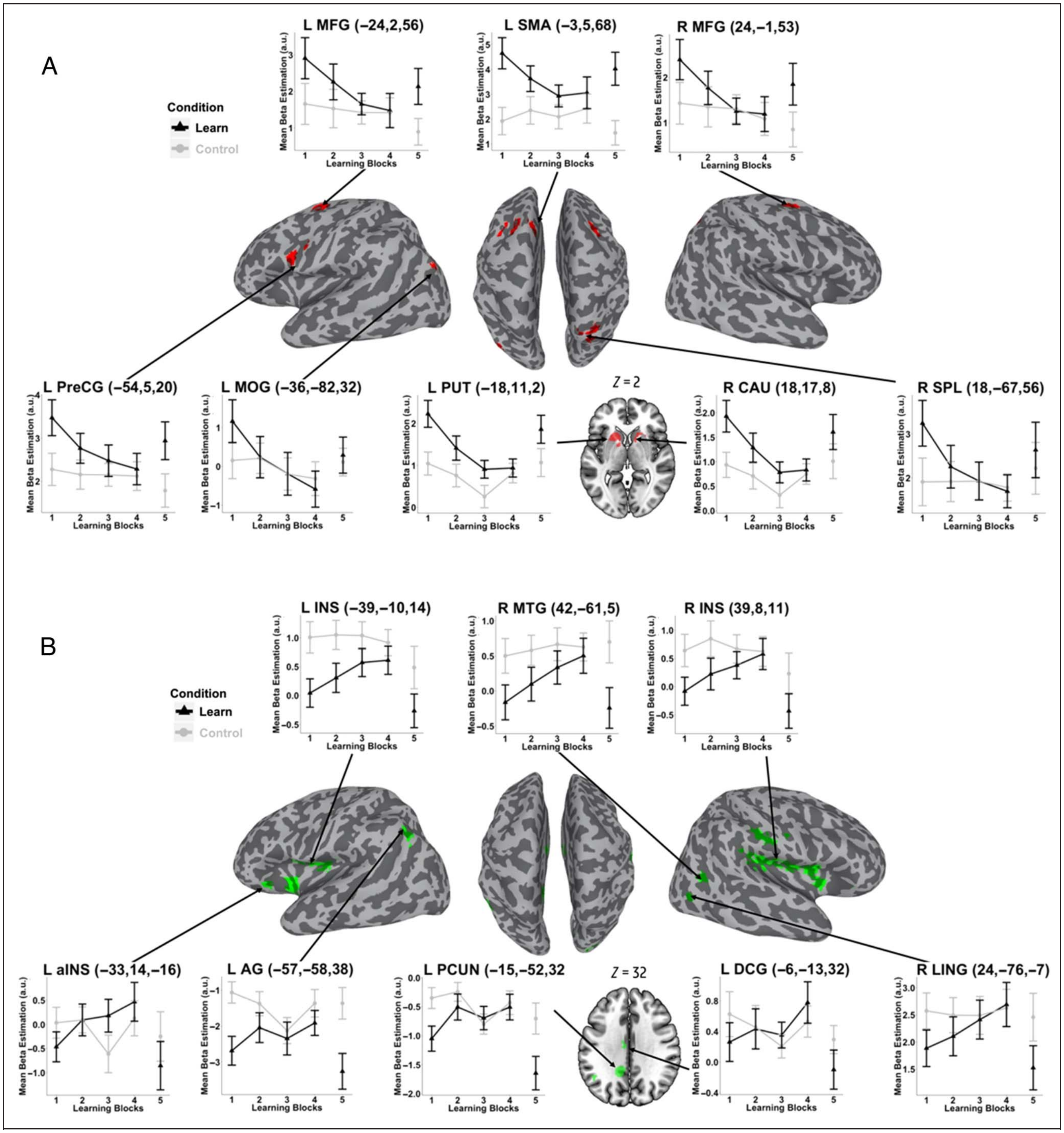

Figure 4. (A) Procedural network. Regions that were involved in counting during the alphabet–arithmetic task (red). (B) Retrieval network. Regions that were involved in retrieving during the alphabet–arithmetic task (green). The average activity profiles are plotted for each region in the procedural and retrieval networks as a function of learning blocks (1 through 5) and learning condition (learn with addends 4 and 7 vs. control with addend 0). The disconnection with Block 5 is to indicate that Block 5 is composed of novel alphabet–arithmetic problems. Error bars represent the standard error of the mean. L = left; R = right; MFG = middle frontal gyrus; PreCG = precentral gyrus; MOG = middle occipital gyrus; PUT = putamen; CAU = caudate; SPL = superior parietal lobe; INS = insula; MTG = middle temporal gyrus; aINS = anterior insula; PCUN = precuneus; DCG = median cingulate gyrus; LING = lingual gyrus.

This cluster-corrected analysis did not reveal significant hippocampal activation, yet, because hippocampal involvement could be expected a priori, we explored the results without imposing a cluster threshold. This may be suggestive of the involvement of the left parahippocamal gyrus (cluster size = 11 voxels, MNI coordinates = −24, −31, −16, $Z = 3.77$) and right HC (cluster size = 27 voxels, MNI coordinates = 42, −16, −10, $Z = 3.79$); see Figure 5.

To evaluate whether activity in these brain networks was related to performance, we correlated a behavioral

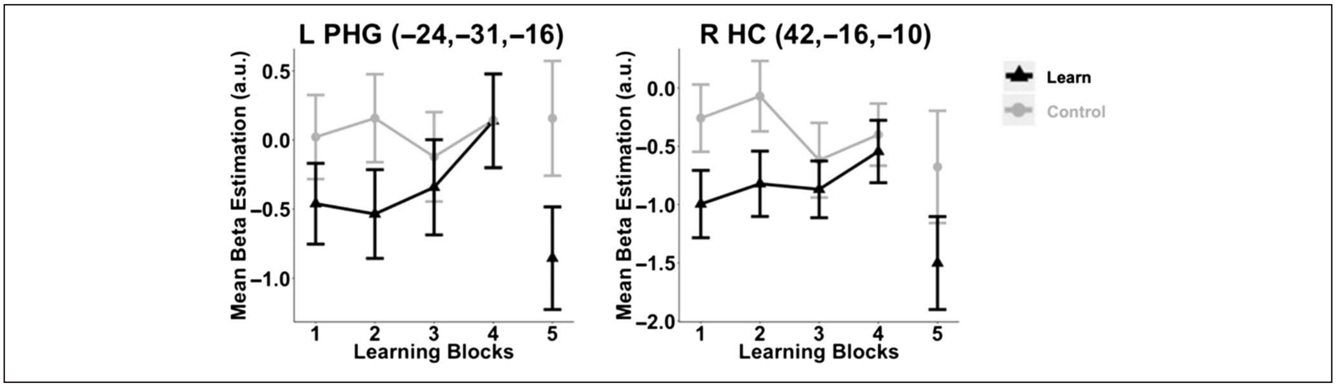

Figure 5. The average activity profiles for the left parahippocampal gyrus (LPH) and R HC as a function of learning blocks (1 through 5) and learning condition (learn with addends 4 and 7 vs. control with addend 0). The disconnection with Block 5 is to indicate that this is a block with novel alphabet–arithmetic problems. Error bars represent the standard error of the mean.

index for learning rate with a neural learning index for each of the two networks. For the *behavioral learning rate index*, we used the Levenberg-Marquard algorithm (Marquardt, 1963; Levenberg, 1944) to estimate the *a* and *k* parameters of the power function $f(x) = ax^{-k}$ of each participant based on the RTs obtained in the learning condition relative to the control condition, with *a* being a general indicator of processing speed and *k* expressing learning rate (average $a = 2823.94$, $SE = 193.37$ and average $k = 1.096$, $SE = .09$).

The *neural learning indices* were computed separately for the learning network and the procedural network following the reasoning behind the whole-brain contrasts. For the procedural network, the neural learning index is obtained as the difference between the beta estimates of Blocks 1 and 5 versus Block 4 in the learning condition and the difference between the beta estimates of Blocks 1 and 5 versus Block 4 in the control condition, that is, as $[(\text{Block1}_{\text{Learn}} + \text{Block5}_{\text{Learn}})/2 - \text{Block4}_{\text{Learn}}] - [(\text{Block1}_{\text{Control}} + \text{Block5}_{\text{Control}})/2 - \text{Block4}_{\text{Control}}]$. The *neural learning index* for the retrieval network is the inverse of the procedural network contrast, that is, $[\text{Block4}_{\text{Learn}} - (\text{Block1}_{\text{Learn}} + \text{Block5}_{\text{Learn}})/2] - [\text{Block4}_{\text{Control}}] - (\text{Block1}_{\text{Control}} + \text{Block5}_{\text{Control}})/2]$. This index was computed for every region of the network and then averaged to obtain a neural learning index for each of the network for every participant. One-sided *t* tests show that the behavioral learning index was positively correlated with the neural learning index of both the procedural ($r = .38$, $t(23) = 1.99$, $p = .029$) and retrieval networks ($r = .46$, $t(23) = 2.46$, $p = .011$) suggesting that the individual differences in learning rate are related to the neural indices of learning. In other words, fast learners displayed stronger learning-related neural changes in both networks (Figure 6).

Effective Connectivity

Effective connectivity analyses were performed to evaluate the learning-related evolution of the strength of the

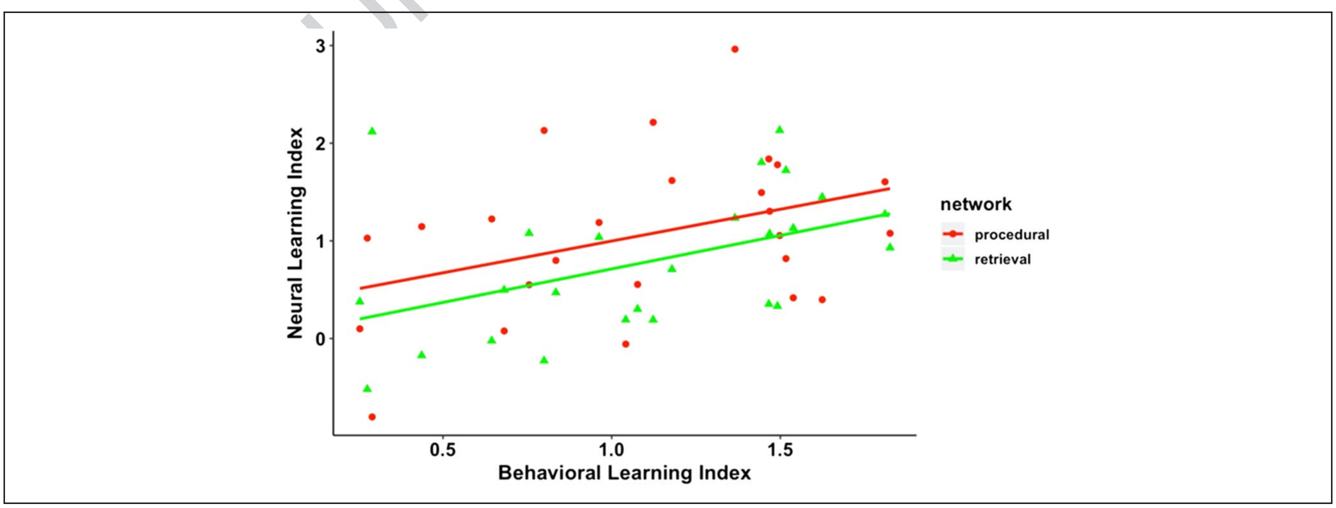

Figure 6. Brain–behavior relationship. The individual differences in learning rate are related to the neural indices of learning. The behavioral learning index is the learning rate estimate from of the of the power curve fitting. The neural learning index is the change in beta estimates over the course of learning. Fast learners displaying stronger learning-related neural changes in both networks. The solid lines represent the linear relationships between the behavioral and neural indices of learning in the procedural and retrieval networks.

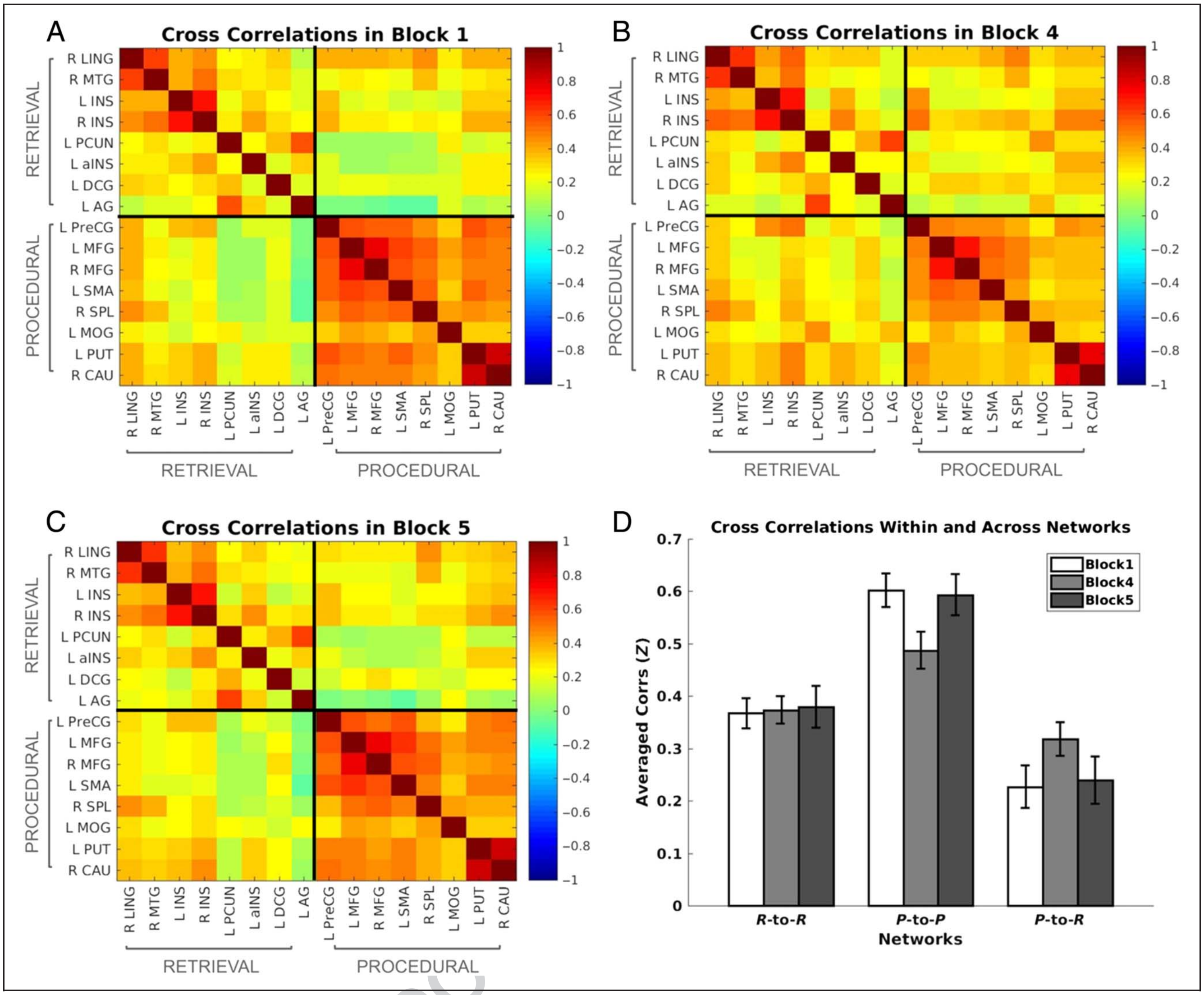

Figure 7. Pairwise connectivity matrices across all the ROIs obtained from the whole-brain analysis in (A) Block 1, (B) Block 4, and (C) Block 5. The connectivity is indexed by the Pearson correlation coefficient, with warm colors representing positive and cold colors negative correlations, averaged across participants. The ROI abbreviations are the same as in Figure 4. (D) Averaged connectivity measures within-network (i.e., *retrieval-to-retrieval* [R-to-R], *procedure-to-procedure* [P-to-P)]) and between-networks (i.e., *procedure-to-retrieval* [P-to-R]) as a function of blocks. Error bars represent the standard error of the mean.

connections within the procedural network (P-to-P connectivity) and retrieval network (R-to-R connectivity) as well as the strength of the connections between these networks (P-to-R connectivity); see Figure 7. The analyses revealed a significant main effect of the network, $F(2, 23) = 98.54$, $p < .001$, $\eta_p^2 = .895$, with stronger within-network connections compared to the between-networks connections (P-to-P > P-to-R, $t(24) = 14.33$, $p < .001$, 95% CI [.25, .34]; and R-to-R > P-to-R, $t(24) = 6.04$, $p < .001$, 95% CI [.07, .15]). The within-network connections themselves were significantly different with P-to-P connections being overall stronger than the R-to-R (P-to-P > R-to-R, $t(24) = 8.51$, $p < .001$, 95% CI [.14, .23]). Although there was no main effect of block ($F < 1$), the connection strengths in the networks varied significantly over the course of the task as indicated by the Block × Network interaction, $F(4, 21) = 10.95$, $p < .001$, $\eta_p^2 = .676$. Post hoc tests revealed that the connections within the procedural network were strong at the beginning of learning, $t(24) = 18.90$, $p < .001$, 95% CI [.53, .67]; then decreased from Block 1 to Block 4, $t(24) = 3.43$, $p < .005$, 95% CI [.046, .18]; and then increased again in the postlearning block, which consisted of learning new problems, $t(24) = -3.022$, $p < .01$, 95% CI [−.17, −.03]. Connections within the retrieval network were significant from the beginning onward, $t(24) = 12.88$, $p < .001$, 95% CI [.31, .43], and remained constant across all learning blocks (no changes of R-to-R strength between blocks; both t values smaller than 1). With respect to the connectivity between the procedural and retrieval network, the functional connectivity analyses show that the two networks are not independent from each other. First, from the beginning onward, both networks are functionally related, $t(24) = 5.6$, $p < .001$, 95% CI [.14, .31]. Second, the strength of the connections *between* networks then increased from Block 1 to Block 4, $t(24) = -3.039$, $p < .01$, 95% CI [−.15, −.02], and decreased again to its original level with the introduction of novel items in the fifth block, $t(24) = 2.53$, $p < .05$, 95% CI [.014, .14]. Some of the correlations between individual regions are negatively signed. However, none of these individual correlations significantly deviated from zero (all $t < 1.35$).

DISCUSSION

This study aimed at unraveling the neural basis of learning alphabet–arithmetic facts, as a proxy of the transition from slow and effortful procedural counting-based processing to fast and effortless processing as it occurs in learning addition arithmetic facts. In particular, we tested two hypotheses. The first hypothesis, the two-systems view, claims that with repeated encounters of the same alphabet–arithmetic problem, a memory trace is formed that after sufficient training allows one to retrieve the answer of the problem from memory. Such memory traces are assumed to be qualitatively different from and rely on different neural sources than the procedural system that is used during initial attempts to solve the problem. The single system hypothesis, on the other hand, argues that fast and effortless processing is not achieved through the development of a qualitatively distinct memory-based system but through a gradual increase in efficiency and speed with which the counting procedures are executed.

The results clearly show the existence of two learning-related neural networks with opposed learning-related changes: one becoming stronger and the other becoming weaker as learning proceeds. This provides strong evidence in support of the two-systems view: Initial learning builds on a procedure-based network, and with learning, activity in this network diminishes, whereas concurrently, a retrieval-based network becomes more active. Interestingly, connectivity analyses show that both networks, despite their opposing trajectory, are functionally connected to each other and that this connectivity in fact increases as learning proceeds.

Behaviorally, the alphabet–arithmetic problems in the first block were solved in a slow and effortful way. A strong distance effect was observed with Distance 7 items being solved hundreds of milliseconds slower than the Distance 4 items, clearly indicating that the counting procedure was involved. The RT increment per unit to be added was 250 msec, corresponding to RT increments in numerical counting (Groen & Parkman, 1972) and in alphabet–arithmetic tasks (Logan & Klapp, 1991). Within the course of the 1-hr session, the addend size effect decreased to a mere 2 msec per unit. The fact that the addend size effect changed so dramatically in such a short time is not easily explained by a gradual speeding up of the counting procedure. It is more in line with a qualitative change in how the problems were solved over repeated exposures. Although the behavioral evidence itself cannot be considered to be conclusive, the neural evidence leaves no room for doubt.

Two neural networks with opposed learning-related changes appeared to subserve these behavioral changes. Activity in a network consisting of basal ganglia and parieto-frontal areas decreased, with learning being in line with a reduction of the involvement of procedure-based processing. Conversely, activity in a network involving the L AG and, to a lesser extent, the HC gradually increases with learning, evidencing the gradual involvement of retrieval-based processing. Such opposed learning-related changes are incompatible with the fast procedures theory, which predicts that changes are restricted to modifications of the procedural counting network that is initially used to solve the new problems. Instead, our results provide direct support for the two-systems view, which assumes learning is based on the formation of memory traces, allowing the fast and effortless retrieval of the solution to known problems, thereby replacing the cumbersome counting procedure.

Assuming that the counting procedures that are executed in the alphabet–arithmetic task are already available in our participants when learning to solve alphabetic problems starts no initial increase in the procedure-related brain network was expected, nor was it observed. In line with the predictions, the decreasing activation as learning of the alphabet–arithmetic problems progresses shows, first, that the counting procedures are invoked in the initial stages of learning and, second, that gradually the contribution of these effortful counting procedures diminishes. The frontoparietal and basal ganglia regions exhibiting such a learning-related decrease correspond to the brain regions that have been associated with the learning and application of procedures, in general, and in the numerical domain in particular. Frontostriatal regions have been found to subserve the initial stages of skill acquisition (e.g., reading mirror-reversed text, Poldrack and Gabrieli [2001] and sequence learning, Gheysen et al. [2011]; Gheysen, Van Opstal, Roggeman, Van Waelvelde, & Fias [2010]). Once the application of the procedures becomes less effortful and automatized, activity in the frontostriatal regions decreases (Poldrack et al., 2005). In the domain of mathematical cognition, learning-related decrease of activity in the frontoparietal network was also observed by Ischebeck et al. (2007) while participants were trained to solve multidigit multiplication facts by drill. Similar to this study, that study found that with increased number of repetitions of the same problem, activity in a strongly overlapping frontostriatal network decreased, in line with a gradually diminishing involvement of procedural processing. Interestingly, the bilateral middle frontal regions, the SMA, posterior parietal regions, and the basal ganglia in our study correspond to a network of brain areas that was shown to be involved more strongly in counting rather than subitizing the numerosity of a set of visually displayed items (Piazza, Mechelli, Butterworth, & Price, 2002), supporting our hypothesis that there are counting procedures that are involved in this procedure network.

Interestingly, the connectivity analysis shows that not only the activity of the procedural network decreased with learning, the strength of the connections within the network also decreased with learning. As learning proceeds and the solution of the alphabet–arithmetic problem relies less on the application of a counting procedure, the neural response of the component regions of the procedure network become less integrated.

As opposed to the decreasing activity in the frontoparietal/basal ganglia network, a number of regions showed learning-related increase of neural activity. These regions comprise the bilateral insular cortex, the middle temporal gyrus and lingual gyrus of the right hemisphere, and the AG in the left hemisphere. The retrieval of known arithmetic facts has repeatedly been shown to activate L AG (Grabner et al., 2007, 2009). Drill training in solving new complex algorithms involving addition and subtraction until the stage of retrieval engages L AG as well (Delazer et al., 2005). Precuneus activation has been shown to co-occur with L AG (Ischebeck et al., 2007; Delazer et al., 2005) and has been taken to be involved in the rapid creation of cortical memory engrams (Brodt et al., 2018). The creation of memory engrams is evidently a key factor on which the alphabet training relies. In this respect, also medial temporal lobe regions like HC can be expected to be related to learning. Our results also point in that direction. Although the medial temporal lobes were not significantly activated when cluster correction was applied, medial temporal lobe (in particular, left parahippocampal cortex and right HC) manifested when an uncorrected threshold was used.

With respect to the connectivity within the retrieval network, the strength of the connections did not change as a function of learning, unlike in the procedural network. A plausible explanation is that the network supporting memory-guided behavior is preconfigured to support memory-guided behavior in general and does not need to be fine-tuned to specific stimuli. So there is no need for the network to adapt to the specific task.

Interestingly, the connectivity between the procedural and the retrieval network sheds light on how the two networks relate to each other. In principle, there are several possibilities. It is possible that from the very beginning, the two networks operate independently from each other, in line with what would be expected from a horse race model according to which the two systems process the incoming information independently to come to a proposed answer, with the final selection being made on the basis of the answer that is proposed first. This possibility is certainly not supported by our results, as the functional connectivity between the procedure and retrieval network is significant from the first block onward. The next question, then, is how this original connection evolves with learning. Does it remain stable, or does it change? And if it changes, does it become weaker, indicating perhaps a winner-take-all learning process in which one solution process, that is, retrieval, dominates the other, that is, procedural, as one could expect based on the univariate results? Interestingly, this is not the case. What happens is that the connectivity between the procedure and retrieval network strengthens. This is remarkable in light of the fact that the activation strength of the procedure network diminishes, as does its internal connectivity. Although a conclusive functional interpretation cannot be provided for this novel finding, it indicates that having learned to efficiently solve alphabet–arithmetic problems does not lead to a solution process that uniquely builds on the retrieval of problem–solution associations, but that the history of procedural counting processes leaves its traces. The procedural involvement may on itself become weaker, but it becomes better integrated with the retrieval network. The fact that the connectivity changes are not permanent but return to their original state when new problems have to be learned suggests that this mechanism of integration is stimulus-specific. Yet, this does not exclude the possibility that with more extensive training, the integration between the counting and the retrieval network could become more structurally integrated, potentially leading to savings in new learning. Further research is needed to test this possibility at a neural level and to verify whether this would be accompanied by a faster behavioral transition from procedure-based to retrieval-based processing when new problems are learned.

An important issue is how our results relate to the neural changes that occur at a developmental time scale. It is clear that similar changes are observed. De Smedt et al. (2011) investigated 10- to 12-year-old children with varying levels of arithmetic skill. Easy small problems that were most likely solved by retrieval activated the left HC and AG, whereas large problems that were most likely solved by (counting) strategies showed stronger frontal activity. Cross-sectional studies have observed a decreasing reliance on frontal areas and basal ganglia as a function of age (from 8 to 19 years; Rivera, Reiss, Eckert, & Menon, 2005) and an increase of activity in AG and medial temporal lobe areas (from second to third grade; Rosenberg-Lee, Barth, & Menon, 2011). Finally, a longitudinal study that investigated the transition from counting to retrieval in 7- to 9-year-old children found a decrease in frontal areas and an increase in HC (Qin et al., 2014). In addition, in terms of connectivity, some noteworthy similarities between our findings and developmental studies have been reported. It has been shown that with development and increase of mathematical skill, the connectivity between HC and frontal and parietal regions including AG increases (Qin et al., 2014; Rosenberg-Lee et al., 2011). In addition, an intervention study showed that third grade children with stronger intrinsic connectivity between HC and frontal regions and basal ganglia responded better to the training than did children with weaker connections (Supekar et al., 2013).

Our study shows that tracking the learning of alphabet–arithmetic problems in a single fMRI session revealed learning-related neural changes on a short time scale that are remarkably similar to the neurodevelopmental trajectory that is found at the much more extended time scale of children learning arithmetic. Of course, one should realize that learning alphabet–arithmetic is only a proxy of development and that limitations have to be taken into account. An important limitation of the current paradigm is that the alphabet–arithmetic problems that were learned were individual and isolated facts, unlike arithmetic problems that are organized in semantic associative networks of interrelated facts (Verguts & Fias, 2005; Stazyk, Ashcraft, & Hamann, 1982).

To conclude, we were able to specify dynamic neural changes during the learning of alphabet–arithmetic facts. Consistent with the two-systems view, we found that learning is subserved by two neural networks, one for procedural counting and one for retrieval. With learning, the strength of the procedural network diminishes, while the strength of the retrieval network increases. Connectivity analyses gave insight in the functional relationship between the two networks. Despite the opposing learning-related trajectories, they become functionally more integrated. Future work is needed to establish how both networks and their interrelation behave after consolidation and how they further evolve with additional practice. Taking alphabet–arithmetic as a proxy for learning arithmetic, the present results can give direction to better understand the learning dynamics of arithmetic fact learning.

Acknowledgments

This research was supported by the following grants awarded to W. F.: FWO/FNRS EOS research project 30446199, GOA BOF UGent BOF17-GOA-004, and FWO Travel Grant V413014N.

Reprint requests should be sent to Wim Fias, Department of Experimental Psychology and Ghent Institute for Functional and Metabolic Imaging, Ghent University, H. Dunantlaan 2, B-9000 Ghent, Belgium, or via e-mail: wim.fias@ugent.be.

Author Contributions

Wim Fias: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Supervision; Writing—Original draft; Writing—Review & editing. Muhammet Ikbal Sahan: Data curation; Formal analysis; Methodology; Writing—Original draft; Writing—Review & editing. Daniel Ansari: Conceptualization; Funding acquisition; Methodology; Project administration; Writing—Review & editing. Ian M. Lyons: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Writing—Review & editing.

Funding Information

Wim Fias, GOA BOF UGent, grant number: BOF17-GOA-004. Wim Fias, FWO Travel Grant, grant number: V413014N. Wim Fias, FWO/FNRS EOS, grant number: 30446199.

Diversity in Citation Practices

A retrospective analysis of the citations in every article published in this journal from 2010 to 2020 has revealed a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .408, W(oman)/M = .335, M/W = .108, and W/W = .149, the comparable proportions for the articles that these authorship teams cited were M/M = .579, W/M = .243, M/W = .102, and W/W = .076 (Fulvio et al., *JoCN*, 33:1, pp. 3–7). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this article report its proportions of citations by gender category to be as follows: M/M = .600, W/M = .150, M/W = .150, and W/W = .100.

REFERENCES

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *Neuroimage*, *26*, 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018, PubMed: 15955494

Barrouillet, P., & Thevenot, C. (2013). On the problem-size effect in small additions: Can we really discard any counting-based account? *Cognition*, *128*, 35–44. https://doi.org/10.1016/j.cognition.2013.02.018, PubMed: 23583543

Brodt, S., Gais, S., Beck, J., Erb, M., Scheffler, K., & Schönauer, M. (2018). Fast track to the neocortex: A memory engram in the posterior parietal cortex. *Science*, *362*, 1045–1048. https://doi.org/10.1126/science.aau2528, PubMed: 30498125

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, *46*, 3–18. https://doi.org/10.1111/j.1469-7610.2004.00374.x, PubMed: 15660640

Cho, S., Ryali, S., Geary, D. C., & Menon, V. (2011). How does a child solve 7 + 8? Decoding brain activity patterns associated with counting and retrieval strategies. *Developmental Science*, *14*, 989–1001. https://doi.org/10.1111/j.1467-7687.2011.01055.x, PubMed: 21884315

Compton, B. J., & Logan, G. D. (1991). The transition from algorithm to retrieval in memory-based theories of automaticity. *Memory & Cognition*, *19*, 151–158. https://doi.org/10.3758/BF03197111, PubMed: 2017038

Delazer, M., Ischebeck, A., Domahs, F., Zamarian, L., Koppelstaetter, F., Siedentopf, C. M., et al. (2005). Learning by strategies and learning by drill—Evidence from an fMRI study. *Neuroimage*, *25*, 838–849. https://doi.org/10.1016/j.neuroimage.2004.12.009, PubMed: 15808984

De Smedt, B., Holloway, I. D., & Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of

arithmetical fluency. *Neuroimage*, *57*, 771–781. https://doi.org/10.1016/j.neuroimage.2010.12.037, PubMed: 21182966
Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine*, *33*, 636–647. https://doi.org/10.1002/mrm.1910330508, PubMed: 7596267
Friston, K. J., Glaser, D. E., Mechelli, A., Turner, R., & Price, C. (2003). Hemodynamic modeling. In R. S. Frackowiak, K. J. Friston, C. Frith, R. J. Dolan, C. Price, S. Zeki, J. Ashburner, & W. D. Penny (Eds.), *Human brain function* (pp. 823–842). Oxford: Academic Press.
Friston, K. J., Holmes, A. P., Price, C. J., Büchel, C., & Worsley, K. J. (1999). Multisubject fMRI studies and conjunction analyses. *Neuroimage*, *10*, 385–396. https://doi.org/10.1006/nimg.1999.0484, PubMed: 10493897
Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2006). A critique of functional localisers. *Neuroimage*, *30*, 1077–1087. https://doi.org/10.1016/j.neuroimage.2005.08.012, PubMed: 16635579
Gheysen, F., Van Opstal, F., Roggeman, C., Van Waelvelde, H., & Fias, W. (2010). Hippocampal contribution to early and later stages of implicit motor sequence learning. *Experimental Brain Research*, *202*, 795–807. https://doi.org/10.1007/s00221-010-2186-6, PubMed: 20195849
Gheysen, F., Van Opstal, F., Roggeman, C., Van Waelvelde, H., & Fias, W. (2011). The neural basis of implicit perceptual sequence learning. *Frontiers in Human Neuroscience*, *5*, 137. https://doi.org/10.3389/fnhum.2011.00137, PubMed: 22087090
Gitelman, D. R., Penny, W. D., Ashburner, J., & Friston, K. J. (2003). Modeling regional and psychophysiologic interactions in fMRI: The importance of hemodynamic deconvolution. *Neuroimage*, *19*, 200–207. https://doi.org/10.1016/S1053-8119(03)00058-2, PubMed: 12781739
Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., & Neuper, C. (2009). To retrieve or to calculate? Left angular gyrus mediates the retrieval of arithmetic facts during problem solving. *Neuropsychologia*, *47*, 604–608. https://doi.org/10.1016/j.neuropsychologia.2008.10.013, PubMed: 19007800
Grabner, R. H., Ansari, D., Reishofer, G., Stern, E., Ebner, F., & Neuper, C. (2007). Individual differences in mathematical competence predict parietal brain activation during mental calculation. *Neuroimage*, *38*, 346–356. https://doi.org/10.1016/j.neuroimage.2007.07.041, PubMed: 17851092
Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review*, *79*, 329–343. https://doi.org/10.1037/h0032950
Ischebeck, A., Zamarian, L., Egger, K., Schocke, M., & Delazer, M. (2007). Imaging early practice effects in arithmetic. *Neuroimage*, *36*, 993–1003. https://doi.org/10.1016/j.neuroimage.2007.03.051, PubMed: 17490893
Ison, M. J., Quian Quiroga, R., & Fried, I. (2015). Rapid encoding of new memories by individual neurons in the human brain. *Neuron*, *87*, 220–230. https://doi.org/10.1016/j.neuron.2015.06.016, PubMed: 26139375
Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, *124*, 83–97. https://doi.org/10.1037/0096-3445.124.1.83, PubMed: 7897342
Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, *2*, 164–168. https://doi.org/10.1090/qam/10666
Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527. https://doi.org/10.1037/0033-295X.95.4.492
Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 179–195. https://doi.org/10.1037/0278-7393.17.2.179
Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, *11*, 431–441. https://doi.org/10.1137/0111030
Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence. *Nature Reviews Neuroscience*, *6*, 177–190. https://doi.org/10.1038/nrn1626, PubMed: 15711599
Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, *32*, 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550, PubMed: 19400715
Piazza, M., Mechelli, A., Butterworth, B., & Price, C. J. (2002). Are subitizing and counting implemented as separate or functionally overlapping processes? *Neuroimage*, *15*, 435–446. https://doi.org/10.1006/nimg.2001.0980, PubMed: 11798277
Poldrack, R. A., & Gabrieli, J. D. E. (2001). Characterizing the neural mechanisms of skill learning and repetition priming: Evidence from mirror reading. *Brain*, *124*, 67–82. https://doi.org/10.1093/brain/124.1.67, PubMed: 11133788
Poldrack, R. A., Sabb, F. W., Foerde, K., Tom, S. M., Asarnow, R. F., Bookheimer, S. Y., et al. (2005). The neural correlates of motor skill automaticity. *Journal of Neuroscience*, *25*, 5356–5364. https://doi.org/10.1523/JNEUROSCI.3880-04.2005, PubMed: 15930384
Qin, S., Cho, S., Chen, T., Rosenberg-Lee, M., Geary, D. C., & Menon, V. (2014). Hippocampal-neocortical functional reorganization underlies children's cognitive development. *Nature Neuroscience*, *17*, 1263–1269. https://doi.org/10.1038/nn.3788, PubMed: 25129076
Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General*, *126*, 288–311. https://doi.org/10.1037/0096-3445.126.3.288
Rivera, S. M., Reiss, A. L., Eckert, M. A., & Menon, V. (2005). Developmental changes in mental arithmetic: Evidence for increased functional specialization in the left inferior parietal cortex. *Cerebral Cortex*, *15*, 1779–1790. https://doi.org/10.1093/cercor/bhi055, PubMed: 15716474
Rosenberg-Lee, M., Barth, M., & Menon, V. (2011). What difference does a year of schooling make? Maturation of brain response and connectivity between 2nd and 3rd grades during arithmetic problem solving. *Neuroimage*, *57*, 796–808. https://doi.org/10.1016/j.neuroimage.2011.05.013, PubMed: 21620984
Seghier, M. L. (2013). The angular gyrus: Multiple functions and multiple subdivisions. *Neuroscientist*, *19*, 43–61. https://doi.org/10.1177/1073858412440596, PubMed: 22547530
Stazyk, E. H., Ashcraft, M. H., & Hamann, M. S. (1982). A network approach to mental multiplication. *Journal at Experimental Psychology: Learning, Memory, and Cognition*, *8*, 320–335. https://doi.org/10.1037/0278-7393.8.4.320
Supekar, K., Swigart, A. G., Tenison, C., Jolles, D. D., Rosenberg-Lee, M., Fuchs, L., et al. (2013). Neural predictors of individual differences in response to math tutoring in primary-grade school children. *Proceedings of the National Academy of Sciences, U.S.A.*, *110*, 8230–8235. https://doi.org/10.1073/pnas.1222154110, PubMed: 23630286
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated

anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*, 273–289. https://doi.org/10.1006/nimg.2001.0978, PubMed: 11771995

Uittenhove, K., Thevenot, C., & Barrouillet, P. (2016). Fast automated counting procedures in addition problem solving: When are they used and why are they mistaken for retrieval? *Cognition*, *146*, 289–303. https://doi.org/10.1016/j.cognition.2015.10.008, PubMed: 26491834

Verguts, T., & Fias, W. (2005). Interacting neighbors: A connectionist model of retrieval in single-digit multiplication. *Memory & Cognition*, *33*, 1–16. https://doi.org/10.3758/BF03195293, PubMed: 15915789

Zamarian, L., Ischebeck, A., & Delazer, M. (2009). Neuroscience of learning arithmetic—Evidence from brain imaging studies. *Neuroscience and Biobehavioral Reviews*, *33*, 909–925. https://doi.org/10.1016/j.neubiorev.2009.03.005, PubMed: 19428500

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

During the preparation of your manuscript, the questions listed below arose. Kindly supply the necessary information.

1. Please verify the change made in the article title.
2. Please confirm if 1deg should be changed to 1°.
3. Please define TRs.
4. Please spell out both occurrences of TE.
5. Please spell out TI.
6. Please define EPI.
7. Please verify the changes made here: [in-plane resolution = 3.3 × 3.3 mm]
8. Please verify the changes made here: As a result of these steps, processed time series associated to learning were obtained for each subject (25), each ROI (16), and each Block (3: Block 1, Block 4, Block 5) separately.
9. Please complete as a sentence if applicable: Fast learners displaying stronger learning-related neural changes in both networks.
10. Please verify the change made here: Assuming that the counting procedures that are executed in the alphabet–arithmetic task are already available in our participants when learning to solve alphabetic problems starts no initial increase in the procedure-related brain network was expected, nor was it observed.
11. Please verify the changes made here: Interestingly, the bilateral middle frontal regions, the SMA, posterior parietal regions, and the basal ganglia in our study correspond to a network of brain areas that was shown to be involved more strongly in counting rather than subitizing the numerosity of a set of visually displayed items (Piazza, Mechelli, Butterworth, & Price, 2002), supporting our hypothesis that there are counting procedures that are involved in this procedure network.
12. Please confirm if the data captured in the Author Contributions and Funding Information sections are complete and accurate.

END OF ALL QUERIES