Improving with Practice: A Neural Model of Mathematical Development

Sean Aubin (saubin@uwaterloo.ca)
Aaron R. Voelker (arvoelke@uwaterloo.ca)
Chris Eliasmith (celiasmith@uwaterloo.ca)
Centre for Theoretical Neuroscience, University of Waterloo
Waterloo, ON, Canada N2L 3G1

Abstract

The ability to improve in speed and accuracy as a result of repeating some task is an important hallmark of intelligent biological systems. Although gradual behavioural improvements from practice have been modelled in spiking neural networks, few such models have attempted to explain cognitive development of a task as complex as addition. In this work, we model the progression from a counting-based strategy for addition to a recall-based strategy. The model consists of two networks working in parallel: a slower basal ganglia loop, and a faster cortical network. The slow network methodically computes the count from one digit given another, corresponding to the addition of two digits, while the fast network gradually "memorizes" the output from the slow network. The faster network eventually learns how to add the same digits that initially drove the behaviour of the slower network. Performance of this model is demonstrated by simulating a fully spiking neural network that includes basal ganglia, thalamus and various cortical areas. Consequently, the model incorporates various neuroanatomical data, in terms of brain areas used for calculation and makes psychologically testable predictions related to frequency of rehearsal. Furthermore, the model replicates developmental progression through addition strategies in terms of reaction times and accuracy, and naturally explains observed symptoms of dyscalculia.

Keywords: neural engineering framework; semantic pointer architecture; nengo; cognitive modelling; mathematical ability; dyscalculia; skill consolidation

Introduction

Adaptability and scalability are two central challenges in cognitive modelling. Building a model that performs some behaviour is a daunting enough task, even without considering how it might improve over time. This is made significantly more challenging when building a large-scale biologically plausible model that uses spiking neurons to represent and communicate information over time.

One example of such a biologically plausible model is the Semantic Pointer Architecture Unified Network (Spaun), which is currently the largest behaving model of the human brain (Eliasmith et al., 2012). This model flexibly performs eight different cognitive tasks by receiving images of handwritten digits through its simulated retina and outputting responses by drawing using its simulated arm. This model makes some progress in addressing the challenge of scalability (Eliasmith, 2013), but lacks the ability to learn from previously experienced cognitive tasks to permanently improve its performance (Spaun only changes its long term connection weights during a simple reinforcement learning task). In this work, we show that one task from Spaun's repertoire, *addition-by-counting*, can be extended to benefit from this cognitive ability.

For addition-by-counting, the model is presented with two digits and asked to draw the digit that corresponds to the sum of these two digits. The "counting" strategy methodically computes the result by adding one to a digit, for the number of times indicated by the other digit. In Spaun, this is accomplished by a number of structures that function together. Visual cortical areas compress the representation of a given image into a semantic representation, referred to as a *semantic pointer*. Prefrontal cortical areas transform these pointers, while maintaining partial results in working memory, and a basal ganglia and thalamus control loop selects actions to coordinate and drive the behaviour of the system. The cortical areas of Spaun should, but are currently unable, to recognize and learn from previous instances of this problem. Including such learning should translate to performance improvements in both speed and accuracy.

DeWolf & Eliasmith (2013) have presented a neural model in which a simple motor skill is consolidated into cortex via repeated practice. However, the skill being learned in that case cannot be extended to solve the addition task because it is unclear how a simple motor action could be generalized to counting from any digit to any other. Consequently, here we consider how consolidation can happen for more sophisticated representations, and argue that such representations may help to explain more complex cognitive phenomena. This is demonstrated by proposing a spiking neural model that exhibits gradual performance improvement on the addition-by-counting task.

A previous model of this task has been implemented in ACT-R, however this model has no neuroanatomical mapping and does not address the problem of transitioning between strategies (Lebiere, 1999).

Mathematical Development

The development of numeracy in children is not a simple progression of skills. For instance, children learn to count before they understand how sets relate to numbers. As well, it is thought that after learning to count, they progressively learn the relationship between set sizes and numbers until the age of five, after which they have a complete understanding of the number scale (Sarnecka & Carey, 2008). This same sort of complex learning development can be observed in children learning addition. Typically, children progress through various strategies before finally memorizing the results of addition, as shown in Table 1 (Siegler, 1987).

The *Counting* strategy involves choosing the larger number and incrementing it a number of times equal to the smaller

Table 1: Percentage of addition strategy use by grade level (Summarized from Siegler (1987)).

Grade level	Counting	Recall	Guess or no response	Other
Kindergarten	30 %	16 %	30 %	24 %
Grade 1	38 %	44 %	8 %	10 %
Grade 2	40 %	45 %	5 %	11 %

Table 2: Median solution times (seconds) per addition strategy use by grade level (Summarized from Siegler (1987)).

Grade level	Counting	Recall
Kindergarten	6.0 s	3.9 s
Grade 1	6.9 s	2.1 s
Grade 2	3.9 s	1.8 s

Table 3: Percentage of errors per addition strategy use by grade level (Summarized from Siegler (1987)).

Grade level	Counting	Recall
Kindergarten	19 %	29 %
Grade 1	4 %	17 %
Grade 2	3 %	7 %

number. The *Recall* strategy is where the two numbers form an association to a previously memorized answer, which is then recalled from long-term memory. Other strategies have been identified, although we focus here on recall and counting, since the greatest developmental change is seen with these two strategies. Additionally, for the sake of simplicity, we focus on the progression from the counting strategy to recall, using sums less than ten. As demonstrated below, the model is initially entirely reliant on the counting strategy, but gradually learns the recall strategy, causing an increase in accuracy and a decrease in reaction times, consistent with the data shown in Table 2 and Table 3.

To this point we have characterized the typical developmental path for children. However, there are individuals who suffer from dyscalculia. Dyscalculia is a learning disability characterized by various problems with numeracy, one of which can be understood as difficulty making the transition from counting to retrieval. Our model demonstrates how, without a parallel learning mechanism, symptoms of dyscalculia can arise. In particular, our model explains increased activation of the prefrontal cortex compared to individuals with normal numeracy (Kucian & von Aster, 2015) and a lack of progression to recall based strategies.

Neural Representation of Digit Semantics

To build our model, we make use of the Neural Engineering Framework (NEF; Eliasmith & Anderson (2003)) and the Semantic Pointer Architecture (SPA; Eliasmith (2013)) that were both used to build Spaun.

The NEF may be understood as a "neural compiler" for mathematical functions using vector spaces. The presented model relies on two key principles of the NEF as a method for constructing networks of spiking neurons and connection weights from a mathematical description of the system. The first principle describes how a vector can be mapped onto a distributed representation over neurons. The second principle characterizes how connections between neurons can transform these vectors.

A vector $\mathbf{x}(t)$ is represented by encoding it into the spiking activity of a population of neurons. Each neuron i has an encoding vector $\mathbf{e}_i$ (which can be understood as a preferred direction in the vector space), a gain α_i, and a background current J_i^{bias}. These parameters determine how the input vector is translated into the input current $J_i(t)$ to a neural nonlinearity $G_i[\cdot]$. For our work, this neural nonlinearity is the leaky integrate-and-fire (LIF) model, which converts the input current into a neural spike train $a_i(t)$.

$$a_i(t) = G_i[J_i(t)], \;\; J_i(t) = \alpha_i \mathbf{e}_i \cdot \mathbf{x}(t) + J_i^{bias} \tag{1}$$

To decode an approximation of the vector back from these spike trains, they are first convolved with a low-pass filter $h(t)$ (a decaying exponential modelled after the postsynaptic current) and then multiplied by a decoding vector $\mathbf{d}_i$.

$$\hat{\mathbf{x}}(t) = \sum_i \mathbf{d}_i \, (a_i * h)(t) \tag{2}$$

The decoders $\mathbf{d}_i$ are found using regularized least squares optimization to minimize the error over the range of inputs $\mathbf{x}$:

$$\int (\mathbf{x} - \hat{\mathbf{x}})^2 \, d\mathbf{x}. \tag{3}$$

To describe how two neural ensembles are connected, we define a weight matrix as the outer product of the encoders and decoders $\omega_{ij} = \mathbf{e}_i \bigotimes \mathbf{d}_j$. The second principle then shows that the input current to neuron i may be rewritten as a weighted summation of its postsynaptic potentials, allowing for the connection of multiple populations of neurons:

$$J_i(t) = \sum_j \alpha_i \, \omega_{ij} \, (a_j * h)(t) + J_i^{bias}. \tag{4}$$

To apply arbitrary transformations to the vector using these connections, we can minimize the decoding error for the desired function:

$$\int (f(\mathbf{x}) - \hat{f}(\mathbf{x}))^2 \, d\mathbf{x}. \tag{5}$$

Together, these two principles allow us to construct spiking neural models of arbitrary functions of vector spaces.

To apply the NEF to cognitive models, the SPA suggests specific architectural components and organization as well as a general kind of representation called a semantic pointer. Semantic pointers are compressed neural representations that can be efficiently manipulated, transformed, and dereferenced to retrieve deep semantic information. It has been suggested that this approach addresses the symbol grounding problem by defining the semantics of a neural representation as resulting from the compression of sensory information as well as conceptual relations (Eliasmith, 2013). Consequently, the SPA lends itself especially well to representing concepts which are grounded in multiple modalities but unified in a single representation. In this model, the semantic pointers are used to represent digits, which have been shown to be grounded in auditory and visual modalities (Nieder, 2012), but also include conceptual relationships (e.g. TWO comes after ONE but before THREE). Conceptual relationships are captured in the SPA by adopting a compression operator from a specific type of Vector Symbolic Architecture (Gayler, 2004) called a Holographic Reduced Representation (Plate, 1995).

Modelling the Counting Strategy

Although Spaun serves as the starting point for our model, it is not practical to simulate all 2.5 million of its neurons. Instead, a counting circuit based on Spaun's design was implemented. As shown in Figure 1, after a question input is received, the procedure consists of three main steps:

1. The digit contained in the working memory neural population is routed to a "transformation system".

2. The digit is transformed by a heteroassociative "incrementing memory" to produce the semantic pointer corresponding to the incremented digit.

3. The incremented number is returned to working memory.

This process is continued until the "Counts finished" are equal to the "Total counts to take". At which point the value in "Count result" is routed as the final answer to the output.

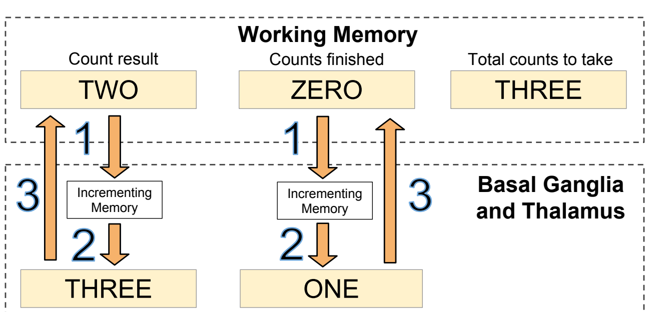

Figure 1: High-level overview of the addition-by-counting procedure. See text for details.

These steps are controlled by an action selection system implemented in the basal ganglia and thalamus (Stewart et al., 2010) in a manner similar to Spaun.

To increment a digit, a predefined heteroassociative memory is used to associate each semantic pointer with its incremented pointer. This is implemented using a single population of neurons, with encoders tuned to each vector of the numerical vocabulary and decoders chosen to apply a transform to the output to compute the incremented semantic pointer. Finally, the output layer uses mutual inhibition to form a winner-take-all (WTA) mechanism, which ensures the transformation system returns only a single pointer. This design is different from Spaun's, which relied on the repeated convolution of a single vector to indicate a count. While both approaches are viable, an associative memory is more effective for simpler, low-dimensional models.

When using a heteroassociative memory, the semantic pointer corresponding to each digit is taken from a random ten-dimensional orthonormal basis. This ensures that no prior information is given to bias the relationship between the ten digits. Initially, the model only knows how to increment a digit, as is the situation for learning addition during childhood using solely the counting strategy.

The learned heteroassociative memory is identical to the aforementioned predefined memory, except the decoders can be learned with experience, so that no action selection mechanism is required to cycle through mappings until the desired result is reached.

Memorization via Reinforcement Learning

The counting circuit that employs the predefined memory is slow, since information is routed back and forth at the rate of subvocal rehearsal. We thus refer to the counting portion of the model as the "Slow-Net". In the model, other cortical areas consolidate the function of the Slow-Net by memorizing its eventual responses. The portion of the network responsible for the storage and subsequent retrieval of originally counted answers will be referred to as the "Fast-Net".

The purpose of the Fast-Net is to learn to associate the two digits provided as input to the Slow-Net with the Slow-Net's eventual response. This heteroassociative memory is highly nonlinear with respect to the vectors that represent each digit. Knight et al. (2016) has shown that by combining the supervised Prescribed Error Sensitivity (PES; MacNeil & Eliasmith (2011)) learning rule with the unsupervised Vector Oja (Voja; Voelker et al. (2014)) learning rule, we can scalably and efficiently learn such complex functions in the NEF. The Fast-Net leverages the unsupervised learning rule to efficiently represent the incoming addends and uses the supervised learning rule to learn the correct sum from the Slow-Net output.

Specifically, the PES learning rule minimizes the difference between the output of a neural population and its desired value, by adjusting its decoders ($\mathbf{d}_i$ from (2)) in response to an error signal. However, for discontinuous high-dimensional

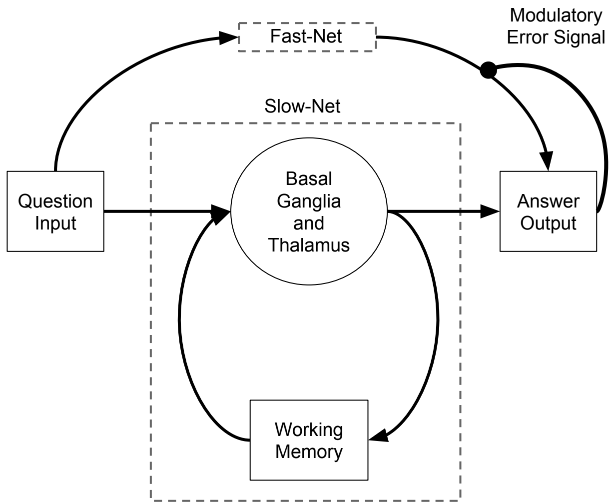

Figure 2: A high-level view of the model, featuring the parallel Slow-Net and Fast-Net.

functions such as the desired heteroassociative memory, this supervised learning rule is insufficient. This is because a neuron will often fire in response to multiple inputs, in which case its decoder will be adjusted to completely overwrite the past association. To avoid this, a neuron should only fire for a single input, which is achievable by selecting the encoders ($\mathbf{e}_i$ from (1)) to be equal to the semantic pointers for each of the digits ahead of time, as is done in the predefined heteroassociative memory. However, this would assume that the area of cortex designated to learn the addition task is already aware of the possible inputs to the Slow-Net.

To keep our approach general, the Voja rule learns to form a sparse encoding of the possible inputs as they are presented to the Fast-Net. This is achieved by adjusting the encoders of any active neurons to become selective to only the current input. This allows PES to learn the correct output without overwriting past associations, while being able to scale to recall over 100,000 associations (Knight et al., 2016).

In our model, to enable simultaneous learning and operation, the Fast-Net is placed in parallel to the Slow-Net, as shown in Figure 2. Both networks receive the same input, but the answer from the Slow-Net is projected to modulate the output of the Fast-Net. This *modulatory error signal* is triggered whenever the Slow-Net responds with an answer, which then provides the feedback for learning via PES. In summary, the population of neurons in the Fast-Net memory continually learns to represent the input digits by adjusting its encoders, while its decoders adjust to associate its input with the output of the Slow-Net.

The learning rate of the heteroassociative memory can be adjusted to model developmentally plausible learning. At high learning rates the model learns mappings after being shown only a single example and at lower learning rates it gains confidence gradually, covering the spectrum of human variability and demonstrating the versatility of the heteroassociative memory. Given the focus of this paper on increasing the accuracy and speed of reaction, while ignoring realistic amounts of rehearsals required to learn, an artificially high learning rate was chosen for the simulation.

Results

The model was built and simulated using Nengo (Bekolay et al., 2014). The results of the Slow-Net, which implement the counting strategy, are shown in Figure 3.

As expected, the magnitude of error between the answer from Fast-Net and the correct answer decreases with rehearsal, as shown in Figure 4. Each epoch of training consists of 20 randomized example additions with no repetition. Given the fast learning rate of the model, after each epoch there is a significant drop in error (during each subsequent epoch, the model is seeing problems it has encountered before). Note that after the magnitude of error drops below 0.5, the Fast-Net response will drive the model's response instead of Slow-Net, since it has the correct answer. Which network drives the overall response is determined by the basal ganglia. Notably, any decrease of error magnitude past 0.5 reflects an increase in the certainty of the answer.

Additionally, once the Fast-Net becomes confident enough in its responses to over-ride the Slow-Net, the speed of responses becomes faster and more uniform, as shown in Figure 5. The confidence of the response is determined by its similarity to any known number representation (correct or incorrect). If the response is incorrect, environmental feedback will drive the Fast-Net to change its answer. Given the high learning rates, such errors are rapidly corrected. The fact that reaction times regularize with transition to a recall-based strategy matches experiments investigating addition strategies (Siegler, 1987).

Neuroanatomical Mappings and Dyscalculia

As discussed in Spaun's mapping of counting (Eliasmith et al., 2012), parietal areas are more active for stable, learned transformations while prefrontal areas are more active for transient, working memory representations. Given that the Fast-Net responses eventually replace those from the Slow-Net, we would expect that with practice brain activity will move from the prefrontal cortex to frontal-parietal areas.

Although this is quite difficult to measure during the natural development of a child, this hypothesis is supported by the observation that those with dyscalculia show greater activation of the prefrontal cortex compared to individuals with normal numeracy (Kucian & von Aster, 2015). Although this model make no claims about why dyscalculia occurs, given that it is a complicated disability usually accompanied by various comorbidities, it does provide a possible explanation as to why such compensation occurs. Specifically, those with dyscalculia are unable to consolidate the functional role of the prefrontal cortex during the counting task within the frontal-parietal region and must instead rely on their working memory. Given an excessively noisy input, inaccurate

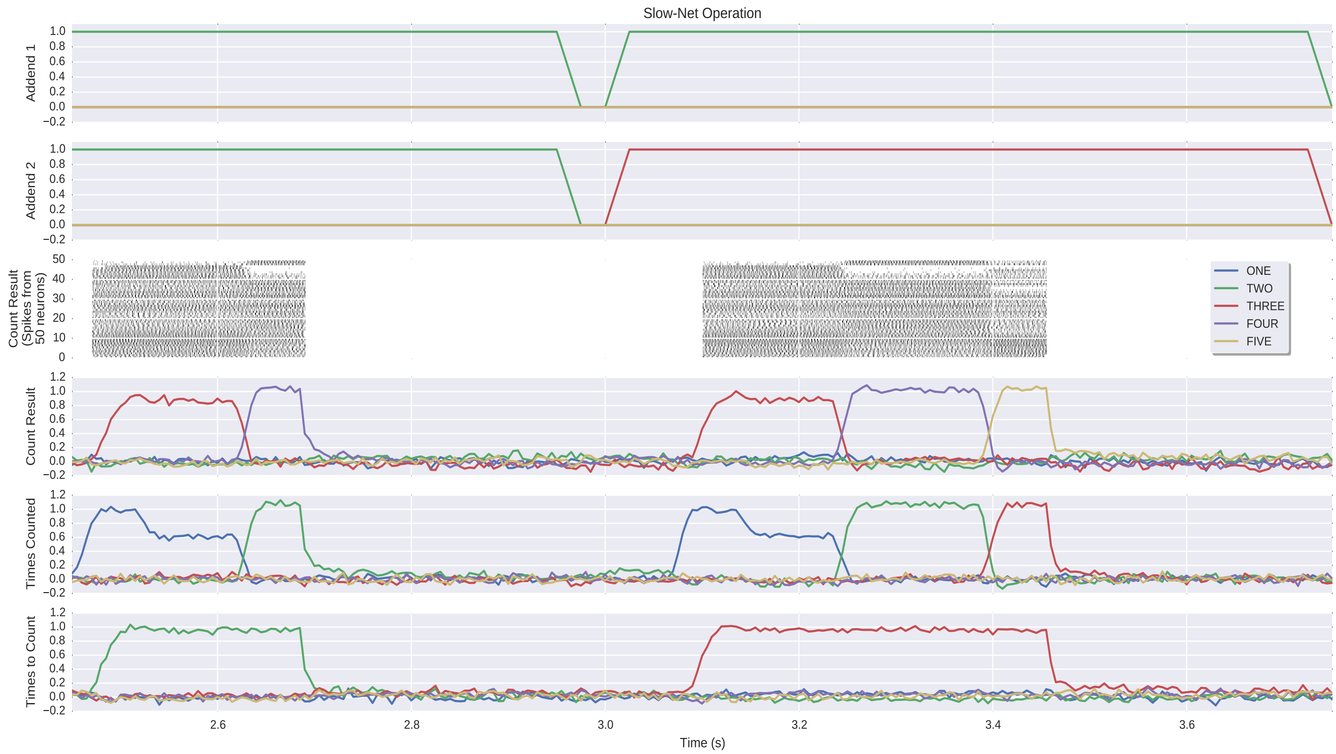

Figure 3: The Slow-Net (counting network) answering the questions $2+2$ and $1+3$. The plots show similarity of neural activity to semantic pointers over time (colours indicate which pointer). As shown in "Count Result" the model is progressing through each intermediate digit before reaching the answer. Lines are the similarity between the spiking pattern in the area and the ideal spiking pattern for each number in the numerical vocabulary.

feedback, or inappropriate modulation of the error signal, the Fast-Net could fail to learn the mapping between addends and sum. Consequently, there would be limited progression from counting to recall and a continued dependence on working memory.

Discussion and Conclusions

We presented a biologically plausible model that progresses from a methodical counting procedure to recalling the response for an addition task. As specified in Table 3 and shown in Figure 4, the accuracy of the Fast-Net memory progresses from noisy to accurate. Once a sufficient accuracy threshold is reached, the memory takes over the process of addition from the Slow-Net, increasing the reaction times.

These simulations suggest the following prediction: that performance should increase in proportion to the frequency of the presented addends. That is, improvement in learning should be proportional to the amount of rehearsal, as well as the familiarity of the given addends. For example, continual presentation of sums with the first addend of "3" should allow for faster learning of other sums involving the addend "3", since the encoders of the heteroassociative memory will have already adapted to represent this digit.

The model as presented is clearly limited in several respects. For instance, numerical representation in the brain consists of more structure than the randomly selected orthonormal vectors used here. This assumption is reasonable for small magnitudes, but is untenable for numerical representation in general. For instance, there is evidence that neurons tuned to numerical size comparisons are proportional to a log scale and can be highly sensitive to task saliency (Nieder & Dehaene, 2009). Capturing such properties will require different numerical representations.

Finally, this model only describes a handful of the brain areas associated with numerical calculation. One of the most surprising areas involved in addition is the cerebellum. It has been suggested that cerebellar activity might be a developmental artifact persisting from when addition is first learned as a physical combination of objects (Arsalidou & Taylor, 2011). To model cerebellar involvement, counting and grouping objects would need to be rehearsed via explicit motor plans. This could be captured with a more in-depth model that includes a motor control hierarchy and a visual system similar to those found in Spaun.

Figure 4: Error magnitude in the Fast-Net decreasing with training received from the feedback of each trial. Note the drop in error after each training epoch.

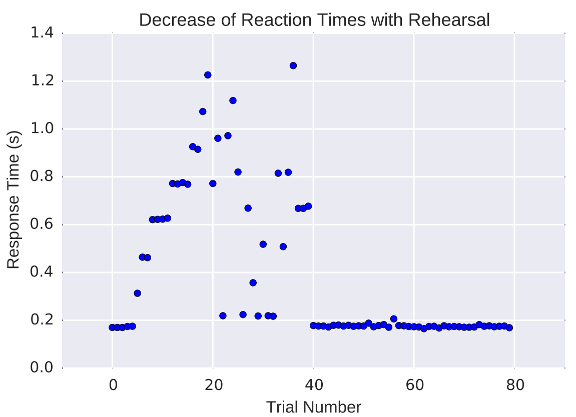

Figure 5: Reaction times decreasing with rehearsal as the Fast-Net takes over for the Slow-Net for increasingly more additions.

Acknowledgments

This work was supported by CFI and OIT infrastructure funding as well as the Canada Research Chairs program, NSERC Discovery grant 261453, AFOSR grant FA8655-13-1-3084 and NSERC graduate funding. The authors are especially grateful to Terry Stewart for his mentorship throughout this work.

References

Arsalidou, M., & Taylor, M. J. (2011). Is 2+ 2= 4? Meta-analyses of brain areas needed for numbers and calculations. *Neuroimage*, *54*(3), 2382–2393.

Bekolay, T., Bergstra, J., Hunsberger, E., DeWolf, T., Stewart, T. C., Rasmussen, D., … Eliasmith, C. (2014). Nengo: A Python tool for building large-scale functional brain models. *Frontiers in Neuroinformatics*, *7*(48). doi: 10.3389/fninf.2013.00048

DeWolf, T., & Eliasmith, C. (2013). A neural model of the development of expertise. In *12th international conference on cognitive modelling.* Ottawa, Ontario.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. New York, NY: Oxford University Press.

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. Cambridge, MA: MIT Press.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science*, *338*, 1202-1205. doi: 10.1126/science.1225266

Gayler, R. W. (2004). Vector symbolic architectures answer Jackendoff's challenges for cognitive neuroscience. *arXiv preprint cs/0412059*.

Knight, J., Voelker, A. R., Mundy, A., Eliasmith, C., & Furber, S. (2016). Efficient SpiNNaker simulation of a heteroassociative memory using the Neural Engineering Framework. In *The 2016 international joint conference on neural networks (IJCNN).* IEEE.

Kucian, K., & von Aster, M. (2015). Developmental dyscalculia. *European journal of pediatrics*, *174*(1), 1–13.

Lebiere, C. (1999). The dynamics of cognition: An ACT-R model of cognitive arithmetic. *Kognitionswissenschaft*, *8*(1), 5–19.

MacNeil, D., & Eliasmith, C. (2011). Fine-tuning and the stability of recurrent neural networks. *PloS one*, *6*(9), e22885.

Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proceedings of the National Academy of Sciences*, *109*(29), 11860–11865.

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual review of neuroscience*, *32*, 185–208.

Plate, T. A. (1995). Holographic reduced representations. *Neural networks, IEEE transactions on*, *6*(3), 623–641.

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, *108*(3), 662–674.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*(3), 250.

Stewart, T. C., Choo, X., & Eliasmith, C. (2010). Symbolic reasoning in spiking neurons: A model of the cortex/basal ganglia/thalamus loop. In *32nd annual conference of the cognitive science society.*

Voelker, A. R., Crawford, E., & Eliasmith, C. (2014). Learning large-scale heteroassociative memories in spiking neurons. In *13th international conference on unconventional computation and natural computation.* London, Ontario.