Limited Neural Capacity and Hyper-Excitability Affect Quantity Processing: A Computational Account

Lani Krossa (LKrossa@scu.edu)
Department of Computer Science and Engineering, Santa Clara University,
500 El Camino Real, Santa Clara, CA 95053, USA

Tannaz Azimi (tazimi@alumni.scu.edu)
Neuroscience Program, Santa Clara University,
500 El Camino Real, Santa Clara, CA 95053, USA

Julia Lieberman (juliaglieberman@gmail.com)
Department of Computer Science and Engineering, Santa Clara University,
500 El Camino Real, Santa Clara, CA 95053, USA

Lang Chen (lchen4@scu.edu)
Department of Psychology, Neuroscience Program, Santa Clara University,
500 El Camino Real, Santa Clara, CA 95053, USA

Abstract

Developmental dyscalculia (DD) is a neurodevelopmental disorder characterized by persistent poor math performance despite normal intelligence and education opportunities. Existing behavioral and neuroimaging studies have demonstrated that quantity processing deficits in DD are accompanied by aberrant brain functions and neurobiological alterations. Although theories have argued that the behavioral impairments observed in DD result from neurobiological deficiency and imbalance of excitatory and inhibitory signals in the brain, these hypotheses are difficult to test in human subjects. Therefore, in the current study, we implemented convolutional neural network models and tested the causal influence of neural capacity (i.e., number of units) and system excitability (i.e., the slope of activation) during the learning of quantity information. For both symbolic and non-symbolic processing, we observed that reducing the number of units did not lead to changes in learning performance. In contrast, increased excitability largely impaired the accuracy of learning, especially for the non-symbolic representations. Therefore, our model simulations provided direct evidence that increased excitability in the brain could result in behavioral impairments in learning quantity information, potentially suggesting a neurobiological basis for DD.

Keywords: developmental dyscalculia; quantity processing; convolutional neural network; neural capacity; system excitability

Introduction

Math skills have been shown to be a predictor of success; math abilities are correlated with mental health (Ashcraft & Ridley, 2005), self-esteem (Kazemi et al., 2014), and pursuing higher education (Bynner, 1997). An estimated 3-6% of children suffer from a neurodevelopmental disorder (Haberstroh & Schulte-Körne, 2019; Reigosa-Crespo et al., 2012; Shalev et al., 2000) known as developmental dyscalculia (DD), characterized by persistent poor math performance despite normal intelligence and education opportunities (Geary, 2011). The behavioral basis of DD has been largely associated with the learning and processing of quantity information (Menon et al., 2021; Piazza et al., 2010; Price et al., 2007), and other domain-general cognitive abilities, such as memory (Cowan et al., 2011; Geary et al., 1991; Murphy et al., 2007), attention (Blair & Razza, 2007; Menon, 2014; Wu et al., 2017), and visuospatial skills (Ashkenazi et al., 2013; Lambert & Spinath, 2018). However, the neurobiological basis of DD remains unclear (Kaufmann et al., 2011; Peters & De Smedt, 2018). Therefore, the current study aims to use computational models to provide insights into how neurobiological factors lead to the cognitive impairments in quantity processing observed in DD.

Often, children with DD face a variety of arithmetic challenges, including difficulty estimating amounts, performing arithmetic operations, and processing quantity overall compared to the typically developing (TD) children (Bulthé et al., 2019; De Smedt et al., 2013; Mussolin et al., 2010). Most critically, individuals with DD have been found to have quantity processing impairments in both non-symbolic and symbolic formats (De Smedt et al., 2013; Gilmore et al., 2010; Kucian et al., 2011; Rousselle & Noël, 2007). Non-symbolic quantity processing refers to the use of concrete objects, e.g., dot arrays, to represent and manipulate quantities, whereas symbolic quantity processing generally involves the use of abstract forms of number knowledge such as Arabic numerals. Mounting studies have shown that quantity processing abilities in both formats in early childhood could predict arithmetic skills in TD individuals (Kolkman et al., 2013; Price & Fuchs, 2016; Schneider et al., 2017). Studies have also shown that DD individuals demonstrate impairments in symbolic

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

processing (Piazza et al., 2010; Schwenk et al., 2017) and maybe non-symbolic processing (Lyons & Ansari, 2015; Salvador et al., 2019; Wong & Chan, 2019). Therefore, one prominent theory argues that the core deficit in DD resides in poor quantity skills (Butterworth, 2011; Dehaene & Wilson, 2007; Price et al., 2007).

Neuroimaging studies have further suggested that the quantity processing deficits in DD are accompanied by reduced brain structures (Jolles et al., 2016; Rotzer et al., 2008) or hypo-activations during math tasks in the intraparietal sulcus (IPS; Price et al., 2007). The anterior IPS is hypothesized to represent and process abstract quantity information (Dehaene et al., 2003; Rosenberg-Lee et al., 2011; Schel & Klingberg, 2016), and numerous studies have shown that it is activated during mental arithmetic, quantity comparison, and number line estimation (Chen et al., 2021; Kucian et al., 2006; Price et al., 2007; Rosenberg-Lee et al., 2015). Thus, the neurobiological deficiency in the IPS found in DD may suggest that DD may originate from a limited neural capacity (Butterworth, 2011; Dehaene & Wilson, 2007; Price et al., 2007).

The behavioral impairment in math processing in DD is also associated with aberrant activations and connectivity patterns, largely hyperactive, in various regions outside the IPS compared to their TD peers (Bulthé et al., 2019; Menon & Chang, 2021; Peters & De Smedt, 2018; Rosenberg-Lee et al., 2015). These observations are largely consistent with a prominent neurobiological theory for neurodevelopmental disorders such as DD, namely, the excitation/inhibition (E/I) imbalance theory (Foss-Feig et al., 2017; Padmanabhan et al., 2017). The E/I imbalance theory posits that various neurodevelopmental disorders such as autism spectrum disorder (ASD) and DD result from dysfunctions in the neural system's ability to retain a reasonable balance between the activity of excitatory and inhibitory neurons. A recent meta-analysis of neuroimaging studies on individuals with math difficulties indeed suggests that DD can be characterized as both insufficient activations in the IPS and excessive activations in widely-distributed brain regions (Anonymous, under review). Therefore, besides the core deficit in the IPS, DD could also stem from the imbalance of the excitatory and inhibitory signals in the brain necessary for critical math knowledge and skills.

Although neuroimaging studies have provided a rich understanding of the neurobiological basis of DD, the neuroimaging techniques are by and large correlational since direct manipulation of neurobiological dysfunctions is challenging. In addition, observations in human subjects are commonly influenced by various unobserved factors, so the causal link between manipulated variables and outcomes is commonly taken with caution. Therefore, these limitations compromise our ability to evaluate the predictions of the core deficit theory and E/I imbalance theory for DD with human subjects. Thus, the current study aims to use computational models to directly test the predictions of both theories. In the computational models, we could fully control the manipulated variables and observe the corresponding outcomes. Previous research has used computational models to characterize cognitive impairments (Farah & McClelland, 1991; Plaut et al., 1996; Seidenberg & McClelland, 1989), as well as variability in cognitive skills such as word reading (Dilkina et al., 2008, 2010; Zevin & Seidenberg, 2006), object recognition (Chen et al., 2017; Lambon Ralph et al., 2007; Rogers et al., 2004), and sentence processing (MacDonald & Christiansen, 2002). Computational models have also suggested that aberrancies in internal model parameters may underlie reading disabilities (Woollams et al., 2007), memory deficits (Chen & Rogers, 2015; Ueno et al., 2011), and various forms of agnosia (Plaut, 2002).

In sum, we aimed to implement computational models of quantity processing and to directly test the core deficit and E/I imbalance theories of DD. Specifically, following a previous attempt (Dilkina et al., 2008), we manipulated the number of units in each layer of the model to approximate the neural capacity. The underlying assumption is that with a limited number of units, the model would have reduced computational capacity to represent abstract knowledge from complex stimuli, resulting in behavioral impairments in learning quantity information. In addition, we manipulated the slope of the activation function in the model with the assumption that a steep slope would lead to excessive excitation in the processing system while keeping the input signals constant. These proposed simulations would show how changes in certain model parameters could lead to differences in model performance, providing insights into how changes in the neurobiological factors could result in behavioral impairments in DD.

Study 1

In the first study, we implemented a convolutional neural network model trained on non-symbolic information. In this study, we planned to test the influence of two factors, namely, the number of units in each layer and the slope in the activation function. This allowed us to test their influence on the learning of abstract quantity information from the non-symbolic inputs with dot arrays. We predicted that the reduced neural capacity (the number of units) and hyper-excitability (a steeper slope in the activation function) would lead to impaired performance in the model in the form of a shallower learning curve.

Methods

Model architecture. A deep convolutional neural network model with the canonical structure (Maturana & Scherer, 2015; Schroff et al., 2015; Yamins et al., 2014) was created and trained to learn the mapping between images of dots and abstract quantity representation. In the baseline model, there was a grid of 28*28 units in the input layer and three convolutional blocks, each with a sequence of convolution, pooling, and normalization layers. There were 64 units in the first convolutional layer, 128 units in the second convolutional layer, and 256 units in the third convolutional

layer. The output from the last convolutional layer was vectorized and projected into a flatten layer with 20 units. Then, all units from the latter layer were fully connected to the output layer with 10 units. Each unit in the output layer corresponded to an output quantity in a localist manner. In another word, each unit in the output layer represented one quantity value from 0 to 9.

Training stimuli and procedures. The images of dots varied in quantity, size, and location, and were generated by an in-house code (for a sample image, see **Figure 1**). In every single run, we generated 1,024 images randomly and vectorized the images as a 28-by-28 matrix for the input units. Then, the model was trained for 10 epochs with a batch size of 100. The model training was conducted with a learning rate of 0.001 with the Adam optimizer (Kingma & Ba, 2015) to adjust the weights based on the loss function of cross-category entropy. For all variations of the simulation, the rectified linear unit (ReLU) function, $\sigma(x) = \max(0, \alpha*x)$, was used to calculate the activations of each layer based on inputs, in which the α was the slope. In the baseline model, the slope α was set to 1.

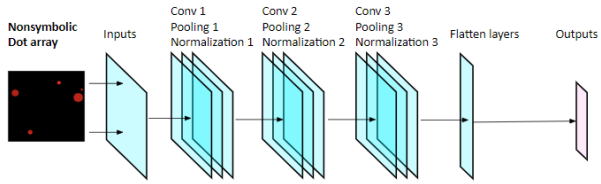

Figure 1: Model architecture of the baseline DNN model for non-symbolic processing with dot arrays

Design and testing. In order to test the effect of neural capacity, we manipulated the number of units proportionally across 3 convolutional layers. In the baseline model, there were 64, 128, and 256 units across the three layers. The number of units in the ½ condition was reduced to 32, 64, and 128, and the number of units in the ⅛ condition was 8, 16, and 32 accordingly. This manipulation was designed to capture the limited neural capacity at the regional level, such as the reduced volume and function of the IPS in the brain. We manipulated another parameter, the slope of the ReLU function, α, to approximate the excitability in the model (**Figure 2**). We used α=0.5 for the hypo-excitability condition and α=2 for the hyper-excitability conditions when the α=1 was set in the balanced baseline model.

These parameters were manipulated in a factorial way (3*3 design), but we were only interested in the main effects of each factor, and their implications to the core deficit and E/I imbalance theories. The learning performance was assessed using the model's accuracy on the training set after each epoch. For each combination of the two factors (a total of 9 combinations), 9 individual runs were conducted with a random set of initial weights. When examining whether each manipulation led to differences in learning, we first conducted a one-way ANOVA with the 3 levels of the parameters (either the number of units or slope) as the between-subject factor on the accuracy after 10 epochs of training. Then, we examined the influence of both parameters across 10 epochs in a mixed-effects ANOVA with each parameter (the number of units or slope) as a between-subject factor and the epoch as a within-subject factor.

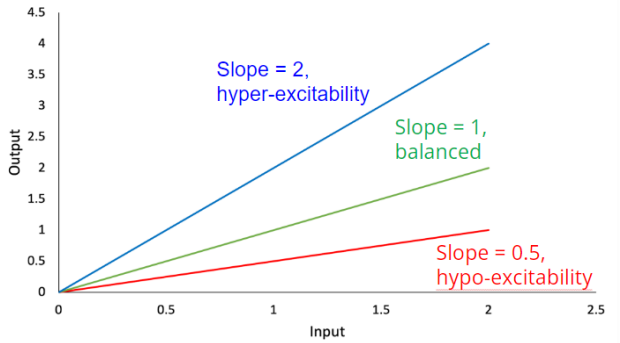

Figure 2: Demonstration of the impact of slope in the ReLU activation function.

Results and Discussion

After 10 training epochs, most of the runs were able to correctly learn the mapping between the images of dot arrays and the corresponding quantity label (**Table 1**). The overall accuracy of all individual runs ranged from 0.82 to 0.96. When examining the main effect of the number of units after training (1, ½, or ⅛ to the baseline model), to our surprise, we observed no main effect of the number of units, $F(2,24) = 0.874$, $p = 0.43$, partial $\eta^2 = 0.068$. Similarly, we also failed to observe a significant main effect of the number of units, $F(2,24) = 0.852$, $p = 0.439$, partial $\eta^2 = 0.066$, or a significant interaction between the parameter and epoch, $F(18,216) = 1.674$, $p = 0.151$ (Greenhouse-Geiser corrected), partial $\eta^2 = 0.122$. These results implied that models with different numbers of units in each layer showed similar learning trajectories over the training (**Figure 3**).

Table 1: Accuracies after 10 epochs of training.

# of Units	Slope	Mean	SD
8	1	0.93	0.028
32	1	0.96	0.035
64	1	0.92	0.115
64	0.5	0.93	0.054
64	1	0.92	0.115
64	2	0.82	0.096

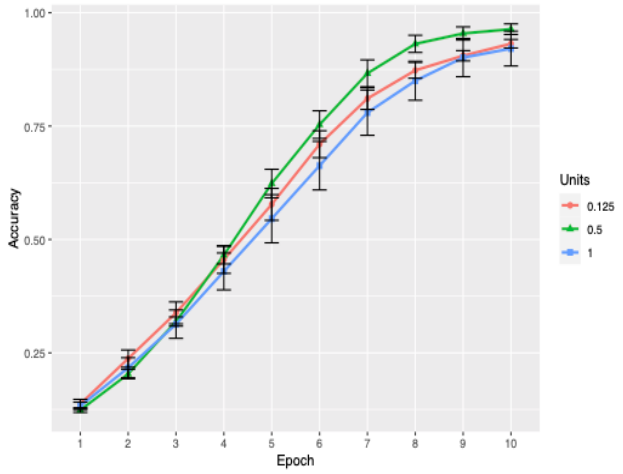

Figure 3: The effect of the number of units on the accuracy of non-symbolic learning. The three levels were ⅛, ½, and 1 proportion of the units in the baseline model.

When examining the main effect of the slope of the activation function (0.5, 1, and 2), we observed a significant main effect after training, $F(2,24) = 4.115$, $p = 0.029$, partial $\eta^2 = 0.255$. Post hoc tests with Tukey-HSD correction revealed that there was a significant difference between when the slope was 0.5 and 2, $p = 0.040$, and a marginally significant difference between when the slope was 1 and 2, $p = 0.067$. These results suggested that when the excitability of the model was excessively high, i.e., hyper-excitatory, the ability to learn quantity information was indeed compromised. When examining the effect of the slope across training epochs, we observed the main effect of the slope, $F(2,24) = 4.049$, $p = 0.031$, partial $\eta^2 = 0.252$. However, there was only a marginally significant slope*epoch interaction, $F(18,216) = 1.936$, $p = 0.099$ (Greenhouse-Geiser corrected), partial $\eta^2 = 0.139$ (**Figure 4**). Overall, the learning performance under the slope conditions of 0.5 and 1 showed little difference, but when the slope was set to 2 in the activation function, the learning was significantly affected compared to the other two conditions.

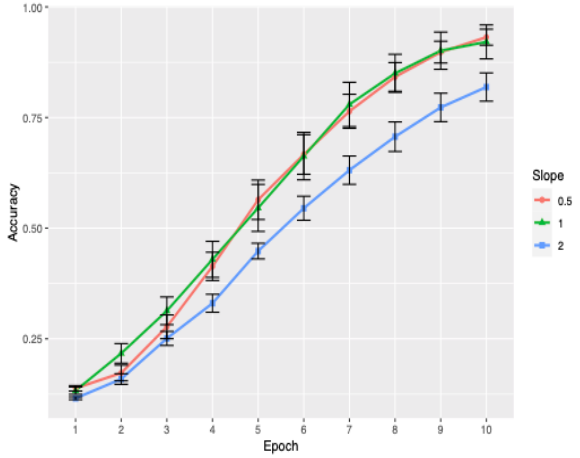

Figure 4: The effect of slope on the accuracy of the non-symbolic learning.

Study 2

In the second study, we implemented a convolutional neural network model trained on symbolic information using images of hand-written digits from the MNIST dataset (LeCun & Cortes, 2010). We aimed to demonstrate the influence of the number of units and the slope in the activation function on the learning of symbolic representations of quantity knowledge. We would like to see if converging findings could be observed for both non-symbolic and symbolic representations of quantity information.

Methods

Model architecture. The same model architecture in Study 1 was adapted here to take the images of hand-written digits as inputs (**Figure 5**).

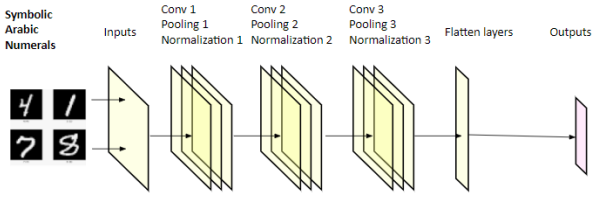

Figure 5: Model architecture of the baseline DNN model for symbolic processing with hand-written digits

Training, design, and testing. The same training, design, and testing procedures as in Study 1 were used here except that the training was based on hand-written images of Arabic digits. In each individual run, we randomly sampled 1,024 images depicting numbers 0 to 9 as the training set as well as the testing set.

Results and Discussion

Similar to the non-symbolic model, after 10 training epochs, most runs could achieve a decent level of learning (**Table 2**). Overall, the learning of the symbolic representations was fast and steep initially, and the final accuracy across different conditions ranged from 0.85 to 0.97.

Table 2: Accuracies after 10 epochs of training.

Units	Slope	Mean	SD
8	1	0.92	0.011
32	1	0.97	0.015
64	1	0.94	0.083
64	0.5	0.95	0.081
64	1	0.94	0.083
64	2	0.85	0.246

The main effect of the number of units (1, ½, and ⅛ of the baseline model) was not significant on the accuracy after 10 epochs, $F(2,24) = 2.366$, $p = 0.115$, partial $\eta^2 = 0.165$. The mixed-effects ANOVA with units and epochs revealed very similar results that there was no significant main effect of the number of units, $F(2,24) = 1.900$, $p = 0.171$, partial $\eta^2 = 0.137$, or interaction between units and epoch, $F(18,216) = 0.519$, $p = 0.718$ (Greenhouse-Geiser corrected), partial $\eta^2 = 0.041$. These results were consistent with the observations in Study 1 for non-symbolic representations (**Figure 6**).

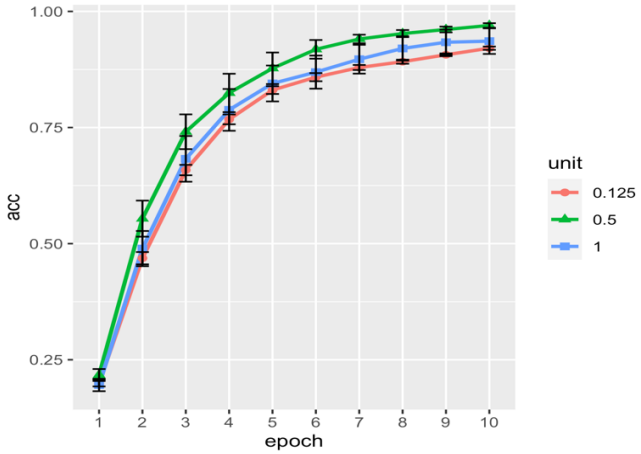

Figure 6: The effect of the number of units on the accuracy of non-symbolic learning. The three levels were ⅛, ½, and 1 proportion of the units in the baseline model.

The results of the slope in the activation function were overall similar to what we observed in Study 1: The condition with a slope of 2 showed the lowest averaged accuracy over learning (**Figure 7**). However, there was a much larger variability across individual runs in this condition. As a result, we failed to observe a significant main effect of slope on the accuracy after training, $F(2,24) = 1.077$, $p = 0.357$, partial $\eta^2 = 0.082$. When examining the effect of the slope across training epochs, we did not observe a significant main effect of the slope, $F(2,24) = 1.242$, $p = 0.307$, partial $\eta^2 = 0.094$, or a significant slope*epoch interaction, $F(18,216) = 0.333$, $p = 0.880$ (Greenhouse-Geiser corrected), partial $\eta^2 = 0.027$. Therefore, we failed to replicate the effect of slope found in Study 1, but the insignificant results were very likely to result from the large variability across individual runs.

General Discussion

Our study implemented a convolutional neural network model to investigate how neural capacity and system excitability may impact learning performance, aiming to provide insights into the neurobiological cause of developmental dyscalculia (DD). Our simulation results showed no evidence on neural capacity (in the form of the

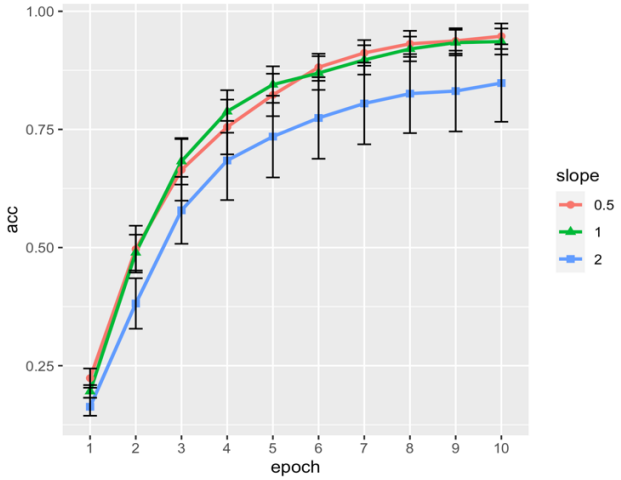

Figure 7: The effect of slope on the accuracy of the symbolic learning.

number of units) impacting learning, but a possible role of over-excitability compromising the learning, especially for the non-symbolic form of quantity knowledge. Thus, our results supported the E/I imbalance theory, positing that the cognitive impairment in DD could result from the excessive excitatory signals in the brain of affected individuals. However, further studies are needed to test the hypothesis of the core deficit theory, arguing that DD results from the neural deficiency in core brain systems, such as the IPS, for quantity and math processing.

Ample neuroimaging studies have shown that individuals with DD are associated with reduced structure or activation in IPS during math processing (Jolles et al., 2016; Price et al., 2007; Rotzer et al., 2008). Therefore, it was surprising to us that reducing the number of units in the models had little impact on learning the quantity information, regardless of the formats (i.e., non-symbolic or symbolic). Previous computational models have shown that limited neural capacity could lead to impaired performance for word reading (Dilkina et al., 2008; Woollams et al., 2007) and object recognition (Chen et al., 2017; Rogers et al., 2004). The reduced number of units in the processing layers could largely decrease the capacity of the models to capture complex and nonlinear mappings between the input and output patterns. Our results seemingly contradict the previous literature on the role of neural capacity in learning. There are a couple of possible reasons for this discrepancy. First, comparing the training materials in previous studies (Chen et al., 2017; Plaut et al., 1996; Rogers et al., 2004), our training stimuli were rather simpler. We only had 1,024 images to map onto 10 quantity labels (0-9). It is possible that the current model, even the one with the least capacity (with only ⅛ of the units in the baseline model), could easily learn the mappings. In other words, learning to map the quantity information was not a challenging task in the 3-layer CNN model with just a few units in our study. Second, recent neuroimaging studies have suggested that the neurobiological deficits in DD may stem from a network of both IPS and

ventral visual pathway (such as the fusiform gyrus) for an integration of non-symbolic and symbolic forms of quantity information (Chen et al., 2021; Fias et al., 2013). Since our model only captured the image processing of quantity information in non-symbolic or symbolic representations separately, it failed to reveal any core deficits that originate from the interaction between multiple representations. Thus, our simulation did not establish the casual link between the core neural deficit (in the form of limited capacity) and cognitive impairments in DD. However, further examinations are needed to show whether the impaired performance on quantity learning could result from reduced neural capacity with more challenging training stimuli and in a more complicated model with multiple representations of different formats (e.g., symbolic and non-symbolic).

The E/I imbalance theory has been proposed to explain various neurodevelopmental disorders (Foss-Feig et al., 2017; Padmanabhan et al., 2017), including DD. Consistent with the E/I imbalance theory, previous studies have indeed reported that individuals with DD showed increased brain activations or functional connectivity in a widely-distributed brain network (Chen et al., 2021; Rosenberg-Lee et al., 2015). A recent meta-analysis revealed that DD is characterized by insufficient neural engagement in core regions for math processing and excessive engagement in other regions for domain-general cognitive skills (Anonymous, under review). Our results then provided direct evidence that over-excitability in the model resulted in impaired quantity learning, consistent with the E/I imbalance theory and previous findings. It is noteworthy that the averaged performance of the hyper-excitability model (i.e., when the slope was set to 2) was numerically lower than the models in the other two conditions with smaller slopes. A couple of reasons could explain the large variability. In our study, for example, we only provided 9 individual runs under each condition. This may be sufficient for the non-symbolic representations but not for the symbolic representations. Another reason is that we randomly selected 1,024 images of hand-written digits as inputs for the symbolic model. Our non-symbolic model was trained on dot arrays with variations in size, location, and quantity, possibly making the training set for each individual run well-matched on the difficulty. However, since the training set in the symbolic model was selected from 60,000 images in the MNIST database, it is very likely that the chosen 1,024 images for the training in each run were not well-matched in visual complexity so that some sets were easier to learn whereas other sets were more challenging. Therefore, more rigorously chosen training materials are needed for future studies to examine the impact of excitability in the model on learning symbolic information. Based on the converging evidence from previous neuroimaging studies and our simulation, we are inclined to suggest that hyper-excitability in the brain could lead to impaired cognitive abilities such as quantity and math skills observed in DD.

Our computational models provided an initial attempt to directly test predictions from the neurobiological theories of DD. However, there are some critical limitations in the current study, and future research should address these limitations to advance our understanding of the neurobiological basis of DD. First, we adopted the canonical architecture of imaging processing for the CNN model with 3-layers. As we have discussed above, the model performance depends on the complexity of the training stimuli and model architecture. Therefore, future studies should run a few testing rounds and choose a model architecture that is optimal for testing the theories. Second, our task in the model was just quantity learning, which is different from what has been used in neuroimaging studies. Future studies should build more complex models that can capture tasks such as quantity comparison in both non-symbolic and symbolic formats (Menon et al., 2021; Piazza et al., 2010; Schneider et al., 2017; Zebian & Ansari, 2012), and test whether similar findings could be observed. Third, future studies should also implement multi-representation or multi-modality structures such as previous models for word reading and object recognition (Chen et al., 2017; Plaut et al., 1996; Ueno et al., 2011). Since DD has been shown to have aberrant functions in various brain regions beyond the IPS (Menon et al., 2021; Menon & Chang, 2021), it is critical for future models to capture the quantity and math processing within a single model that could process both non-symbolic and symbolic representations.

In conclusion, our computational models adopted a theory-driven approach to test the causal effects of neural capacity and E/I imbalance in the model on quantity learning. Our results suggest that the impaired quantity processing observed in individuals with DD is likely to result from atypical patterns of excitatory and inhibitory signals in the brain. However, the role of neural capacity or core neural deficiency in the brain on cognitive impairments in DD awaits future investigation.

References

Ashcraft, M. H., & Ridley, K. S. (2005). Math anxiety and its cognitive consequences: a tutorial review. In *The Handbook of Mathematical Cognition* (pp. 315–327).

Ashkenazi, S., Rosenberg-Lee, M., Metcalfe, A. W. S. S., Swigart, A. G., & Menon, V. (2013). Visuo-spatial working memory is an important source of domain-general vulnerability in the development of arithmetic cognition. *Neuropsychologia*, *51*(11), 2305–2317. https://doi.org/10.1016/j.neuropsychologia.2013.06.031

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647–663. isi:000245099600018

Bulthé, J., Prinsen, J., Vanderauwera, J., Duyck, S., Daniels, N., Gillebert, C. R., Mantini, D., Op de Beeck, H. P., & De Smedt, B. (2019). Multi-method brain imaging reveals impaired representations of number as well as

altered connectivity in adults with dyscalculia. *NeuroImage*, *190*(June 2018), 289–302. https://doi.org/10.1016/j.neuroimage.2018.06.012

Butterworth, B. (2011). Foundational Numerical Capacities and the Origins of Dyscalculia. In *Space, Time and Number in the Brain* (pp. 249–265). Academic Press. https://doi.org/10.1016/B978-0-12-385948-8.00016-5

Bynner, J. M. (1997). Basic skills in adolescents' occupational preparation. *Career Development Quarterly*, *45*(4), 305–321. https://doi.org/10.1002/j.2161-0045.1997.tb00536.x

Chen, L., Iuculano, T., Mistry, P., Nicholas, J., Zhang, Y., & Menon, V. (2021). Linear and nonlinear profiles of weak behavioral and neural differentiation between numerical operations in children with math learning difficulties. *Neuropsychologia*, *160*, 107977. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2021.107977

Chen, L., Lambon Ralph, M. A., & Rogers, T. T. (2017). A unified model of human semantic knowledge and its disorders. *Nature Human Behaviour*, *1*(March), 39. https://doi.org/10.1038/s41562-016-0039

Chen, L., & Rogers, T. T. (2015). A Model of Emergent Category-specific Activation in the Posterior Fusiform Gyrus of Sighted and Congenitally Blind Populations. *Journal of Cognitive Neuroscience*, *27*(10), 1981–1999. https://doi.org/10.1162/jocn_a_00834

Cowan, R., Donlan, C., Shepherd, D. L., Cole-Fletcher, R., Saxton, M., & Hurry, J. (2011). Basic Calculation Proficiency and Mathematics Achievement in Elementary School Children. *Journal of Educational Psychology*, *103*(4), 786–803. https://doi.org/10.1037/a0024556

De Smedt, B., Noël, M. P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. In *Trends in Neuroscience and Education* (Vol. 2, Issue 2, pp. 48–55). Urban & Fischer. https://doi.org/10.1016/j.tine.2013.06.001

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*(3–6), 487–506. https://doi.org/10.1080/02643290244000239

Dehaene, S., & Wilson, A. J. (2007). Number sense and developmental dyscalculia. In *Human behavior, learning, and the developing brain: Atypical development.* (pp. 212–238). https://psycnet.apa.org/record/2007-04220-009

Dilkina, K., McClelland, J. L., & Plaut, D. C. (2008). A single-system account of semantic and lexical deficits in five semantic dementia patients. *Cognitive Neuropsychology*, *25*(2), 136–164. https://doi.org/10.1080/02643290701723948

Dilkina, K., McClelland, J. L., & Plaut, D. C. (2010). Are there mental lexicons? the role of semantics in lexical decision. *Brain Research*, *1365*, 66–81. https://doi.org/10.1016/j.brainres.2010.09.057

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: modality specificity and emergent category specificity. *Journal of Experimental Psychology. General*, *120*(4), 339–357. https://doi.org/10.1037/0096-3445.120.4.339

Fias, W., Menon, V., & Szucs, D. (2013). Multiple components of developmental dyscalculia. *Trends in Neuroscience and Education*, *2*(2), 43–47. https://doi.org/10.1016/j.tine.2013.06.006

Foss-Feig, J. H., Adkinson, B. D., Ji, J. L., Yang, G., Srihari, V. H., McPartland, J. C., Krystal, J. H., Murray, J. D., & Anticevic, A. (2017). Searching for Cross-Diagnostic Convergence: Neural Mechanisms Governing Excitation and Inhibition Balance in Schizophrenia and Autism Spectrum Disorders. *Biological Psychiatry*, *81*(10), 848–861. https://doi.org/10.1016/j.biopsych.2017.03.005

Geary, D. C. (2011). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of Developmental and Behavioral Pediatrics*, *32*(3), 250–263. https://doi.org/10.1097/DBP.0b013e318209edef

Geary, D. C., Brown, S. C., & Samaranayake, V. A. (1991). Cognitive Addition: A Short Longitudinal Study of Strategy Choice and Speed-of-Processing Differences in Normal and Mathematically Disabled Children. *Developmental Psychology*, *27*(5), 787–797. https://doi.org/10.1037/0012-1649.27.5.787

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition*, *115*(3), 394–406. https://doi.org/10.1016/J.COGNITION.2010.02.002

Haberstroh, S., & Schulte-Körne, G. (2019). Diagnostik und Behandlung der Rechenstörung. *Deutsches Arzteblatt International*, *116*(7), 107–114. https://doi.org/10.3238/arztebl.2019.0107

Jolles, D., Ashkenazi, S., Kochalka, J., Evans, T., Richardson, J., Rosenberg-Lee, M., Zhao, H., Supekar, K., Chen, T., & Menon, V. (2016). Parietal hyper-connectivity, aberrant brain organization, and circuit-based biomarkers in children with mathematical disabilities. *Developmental Science*, *19*(4), 613–631. https://doi.org/10.1111/desc.12399

Kaufmann, L., Wood, G., Rubinsten, O., & Henik, A. (2011). Meta-analyses of developmental fMRI studies investigating typical and atypical trajectories of number processing and calculation. *Developmental Neuropsychology*, *36*(6), 763–787. https://doi.org/10.1080/87565641.2010.549884

Kazemi, R., Momeni, S., & Abolghasemi, A. (2014). The Effectiveness of Life Skill Training on Self-esteem and Communication Skills of Students with Dyscalculia. *Procedia - Social and Behavioral Sciences*, *114*, 863–866. https://doi.org/10.1016/j.sbspro.2013.12.798

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. M. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction*, *25*, 95–103. https://doi.org/10.1016/J.LEARNINSTRUC.2012.12.001

Kucian, K., Loenneker, T., Dietrich, T., Dosch, M., Martin, E., & von Aster, M. (2006). Impaired neural networks for approximate calculation in dyscalculic children: a functional MRI study. *Behavioral and Brain Functions*, *2*, 31. https://doi.org/10.1186/1744-9081-2-31

Kucian, K., Loenneker, T., Martin, E., & Von Aster, M. (2011). Non-symbolic numerical distance effect in children with and without developmental dyscalculia: A parametric fMRI study. *Developmental Neuropsychology*, *36*(6), 741–762. https://doi.org/10.1080/87565641.2010.549867

Lambert, K., & Spinath, B. (2018). Conservation Abilities, Visuospatial Skills, and Numerosity Processing Speed: Association With Math Achievement and Math Difficulties in Elementary School Children. *Journal of Learning Disabilities*, *51*(3), 223–235. https://doi.org/10.1177/0022219417690354

Lambon Ralph, M. A., Lowe, C., & Rogers, T. T. (2007). Neural basis of category-specific semantic deficits for living things: Evidence from semantic dementia, HSVE and a neural network model. *Brain*, *130*(4), 1127–1137. https://doi.org/10.1093/brain/awm025

LeCun, Y., & Cortes, C. (2010). *MNIST handwritten digit database*. AT&T Labs [Online]. Available: Http://Yann. Lecun. Com/Exdb/Mnist.

Lyons, I. M., & Ansari, D. (2015). Foundations of Children's Numerical and Mathematical Skills: The Roles of Symbolic and Nonsymbolic Representations of Numerical Magnitude. *Advances in Child Development and Behavior*, *48*, 93–116. https://doi.org/10.1016/BS.ACDB.2014.11.003

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*(1), 35–54. https://doi.org/10.1037/0033-295X.109.1.35

Maturana, D., & Scherer, S. (2015). VoxNet: A 3D Convolutional Neural Network for real-time object recognition. *IEEE International Conference on Intelligent Robots and Systems*, *2015-Decem*, 922–928. https://doi.org/10.1109/IROS.2015.7353481

Menon, V. (2014). Arithmetic in child and adult brain. In K. R. Cohen & A. Dowker (Eds.), *Handbook of mathmatical cognition*. Oxford University Press. https://doi.org/doi:10.1093/oxfordhb/9780199642342.013.041

Menon, V., & Chang, H. (2021). Emerging neurodevelopmental perspectives on mathematical learning. *Developmental Review*, *60*, 100964. https://doi.org/10.1016/j.dr.2021.100964

Menon, V., Padmanabhan, A., & Schwartz, F. (2021). Cognitive neuroscience of dyscalculia and math learning disabilities. In *The Oxford Handbook of Developmental Cognitive Neuroscience*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198827474.013.23

Murphy, M. M., Mazzocco, M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities*, *40*(5), 458–478.

Mussolin, C., De Volder, A., Grandin, C., Schlögel, X., Nassogne, M. C., & Noël, M. P. (2010). Neural correlates of symbolic number comparison in developmental dyscalculia. *Journal of Cognitive Neuroscience*, *22*(5), 860–874. https://doi.org/10.1162/jocn.2009.21237

Padmanabhan, A., Lynch, C. J., Schaer, M., & Menon, V. (2017). The Default Mode Network in Autism. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *2*(6), 476–486. https://doi.org/10.1016/j.bpsc.2017.04.004

Peters, L., & De Smedt, B. (2018). Arithmetic in the developing brain: A review of brain imaging studies. *Developmental Cognitive Neuroscience*, *30*(November 2016), 265–279. https://doi.org/10.1016/j.dcn.2017.05.002

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*(1), 33–41. https://doi.org/10.1016/J.COGNITION.2010.03.012

Plaut, D. C. (2002). Graded modality-specific specialisation in semantics: A computational account of optic aphasia. *Cognitive Neuropsychology*, *19*(7), 603–639.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding Normal and Impaired Word Reading: Computational Principles in {Quasi-Regular} Domains. *Psychological Review*, *103*(1), 56–115. https://doi.org/10.1037/0033-295X.103.1.56

Price, G. R., & Fuchs, L. S. (2016). The Mediating Relation between Symbolic and Nonsymbolic Foundations of Math Competence. *PLOS ONE*, *11*(2), e0148981. https://doi.org/10.1371/JOURNAL.PONE.0148981

Price, G. R., Holloway, I., Räsänen, P., Vesterinen, M., & Ansari, D. (2007). Impaired parietal magnitude processing in developmental dyscalculia. *Current Biology*, *17*(24), R1042–R1043. https://doi.org/10.1016/J.CUB.2007.10.013

Reigosa-Crespo, V., Valdés-Sosa, M., Butterworth, B.,

Estévez, N., Rodríguez, M., Santos, E., Torres, P., Suárez, R., & Lage, A. (2012). Basic numerical capacities and prevalence of developmental dyscalculia: The Havana survey. *Developmental Psychology*, *48*(1), 123–135. https://doi.org/10.1037/a0025356

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). The structure and deterioration of semantic memory: a computational and neuropsychological investigation. *Psychological Review*, *111*(1), 205–235.

Rosenberg-Lee, M., Ashkenazi, S., Chen, T., Young, C. B., Geary, D. C., & Menon, V. (2015). Brain hyper-connectivity and operation-specific deficits during arithmetic problem solving in children with developmental dyscalculia. *Developmental Science*, *18*(3), 351–372. https://doi.org/10.1111/desc.12216

Rosenberg-Lee, M., Chang, T. T., Young, C. B., Wu, S. S., & Menon, V. (2011). Functional dissociations between four basic arithmetic operations in the human posterior parietal cortex: A cytoarchitectonic mapping study. *Neuropsychologia*, *49*(9), 2592–2608. https://doi.org/10.1016/j.neuropsychologia.2011.04.035

Rotzer, S., Kucian, K., Martin, E., Aster, M. von, Klaver, P., & Loenneker, T. (2008). Optimized voxel-based morphometry in children with developmental dyscalculia. *NeuroImage*, *39*(1), 417–422. https://doi.org/10.1016/J.NEUROIMAGE.2007.08.045

Rousselle, L., & Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition*, *102*(3), 361–395. https://doi.org/10.1016/J.COGNITION.2006.01.005

Salvador, L. de S., Moura, R., Wood, G., & Haase, V. G. (2019). Cognitive heterogeneity of math difficulties: A bottom-up classification approach. *Journal of Numerical Cognition*, *5*(1), 55–85. https://doi.org/10.5964/jnc.v5i1.60

Schel, M. A., & Klingberg, T. (2016). Specialization of the Right Intraparietal Sulcus for Processing Mathematics During Development. *Cerebral Cortex*, 1–11. https://doi.org/10.1093/cercor/bhw246

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science*, *20*(3), e12372. https://doi.org/10.1111/desc.12372

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June*, 815–823. https://doi.org/10.1109/CVPR.2015.7298682

Schwenk, C., Sasanguie, D., Kuhn, J. T., Kempe, S., Doebler, P., & Holling, H. (2017). (Non-)symbolic magnitude processing in children with mathematical difficulties: A meta-analysis. *Research in Developmental Disabilities*, *64*, 152–167. https://doi.org/10.1016/J.RIDD.2017.03.003

Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed , Developmental Model of Word Recognition and Naming. *Psychological Review*, *96*(4), 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Shalev, R. S., Auerbach, J., Manor, O., & Gross-Tsur, V. (2000). Developmental dyscalculia: Prevalence and prognosis. *European Child and Adolescent Psychiatry*, *9*(SUPPL. 2), S58–S64. https://doi.org/10.1007/s007870070009

Ueno, T., Saito, S., Rogers, T. T., Lambon Ralph, M. A., Lambon Ralph, M. A., & Lambon Ralph, M. A. (2011). Lichtheim 2: Synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, *72*(2), 385–396. https://doi.org/10.1016/j.neuron.2011.09.013

Wong, T. T. Y., & Chan, W. W. L. (2019). Identifying children with persistent low math achievement: The role of number-magnitude mapping and symbolic numerical processing. *Learning and Instruction*, *60*, 29–40. https://doi.org/10.1016/j.learninstruc.2018.11.006

Woollams, A. M., Lambon Ralph, M. A., Plaut, D. C., & Patterson, K. (2007). SD-squared: On the association between semantic dementia and surface dyslexia. *Psychological Review*, *114*(2), 316–339. anna.woollams@mrc-cbu.cam.ac.uk

Wu, S. S., Chen, L., Battista, C., Smith Watts, A. K., Willcutt, E. G., & Menon, V. (2017). Distinct Influences of Affective and Cognitive Factors on Children's Non-verbal and Verbal Mathematical Abilities. *Cognition*, *In Press*.

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8619–8624. https://doi.org/10.1073/pnas.1403112111

Zebian, S., & Ansari, D. (2012). Differences between literates and illiterates on symbolic but not nonsymbolic numerical magnitude processing. *Psychonomic Bulletin & Review*, *19*(1), 93–100. https://doi.org/10.3758/s13423-011-0175-9

Zevin, J. D., & Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword naming: A comparison of current models. *Journal of Memory and Language*, *54*(2), 145–160. http://www.sciencedirect.com/science/article/B6WK4-4HWXP0F-1/2/71e57a42bbf90a21ad635f56d094115d