Categories from dimensions: Population-level computational modelling of neurodevelopmental conditions

Qidong Song (qsong01@mail.bbk.ac.uk)
Department of Psychological Sciences, Birkbeck, University of London
Malet St, London WC1E 7HX United Kingdom

Michael Thomas (m.thomas@bbk.ac.uk)
Department of Psychological Sciences, Birkbeck, University of London
Malet St, London WC1E 7HX United Kingdom

Abstract

Theoretical understanding of neurodevelopmental conditions (NCs) has shifted from a categorical approach to a dimensional one, characterized by an acceptance of comorbidity and heterogeneity. Previous computational modelling of NCs has tended only to accommodate categorical views. The current work presents a mechanistic simulation framework that fits with the dimensional view, using artificial neural networks to model populations of learners, with underlying causes of variation in developmental outcomes viewed as continuous, polygenic, and in part environmental. We show how the dimensional and categorical approaches can be linked using latent profile analysis and outlier methods, recovering profiles and specific deficits from dimensional variation. We show how altering the distribution of hyper-parameters shifts the population composition of developmental profiles and frequencies of deficit patterns, and we test their robustness to stochastic factors.

Keywords: population-level computational modeling; neurodevelopmental conditions; development; polygenic causes; socioeconomic status

Introduction

All classifications in all sciences make distinctions more exact and abrupt than that exist in nature (John Hughlings Jackson, 1874; reprinted in Jackson, 1958, p.202)

The 19th Century British neurologist, John Hughlings Jackson, who specialized in understanding and diagnosing diverse forms of epilepsy, recognized limitations in seeking to assign mental disorders to particular categories. Historically there has been wide adoption of the categorical approach to diagnosing neurodevelopmental conditions such as autistic spectrum condition (autism) and attention deficit hyperactivity disorder (ADHD). In part, this aligns with the practical requirement of making decisions about intervention and resource allocation. The categorical approach to neurodevelopmental conditions presumes first that conditions will differ from each other in their symptoms, and second that those with a particular condition will be similar to each other. It has also tended to be associated with the view that individuals will only have a single condition. For example, prior to the publication of the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013), an individual could not be diagnosed as having both autism and ADHD.

Amplifying Jackson's early reservations, the categorical approach to neurodevelopmental conditions has been challenged in two ways. Individuals who receive diagnoses for a given condition frequently show heterogeneous cognitive profiles rather than being similar to each other, and comorbidity (meeting diagnostic criteria for more than one condition) appears to be the rule rather than the exception (Gillberg, 2010). For example, one recent paper suggested that 50-70% of individuals with autism also present with comorbid ADHD (Hours et al., 2022). In clinical practice, development also appears to be crucial. Gillberg (2010) argued that for children referred for clinical evaluation when 3-5 years old – presenting with delay or differences in general development, language and communication, social inter-relatedness, motor coordination, attention, activity, behavior, mood, and/or sleep – it is often difficult to distinguish conditions. These only differentiate in mid-childhood into clearer profiles such as autism, ADHD, oppositional defiant disorder, developmental coordination disorder, and intellectual and developmental disabilities.

One response has been to reject the categorical approach in favor of a dimensional one, where conditions are seen as differing in their positions over several dimensions, with each viewed as a continuum. For example, in what they labeled a transdiagnostic approach, Astle et al. (2022) suggested that autism, ADHD, dyslexia, dyscalculia, developmental language disorder (DLD), and social pragmatic communication disorder could be re-conceived as different degrees of atypicality on the dimensions of hyperactivity and impulsivity, inattention, social communication, executive functioning, and phonological processing. This framework both allows for sharing of symptoms across conditions and heterogeneity within them.

The dimensional approach has been supported by data-driven, diagnosis-blind approaches that collect data on cognitive profiles of large groups of children who are struggling learners in one or more domains. Machine learning or statistical methods are used to identify clusters of individuals within the sample exhibiting different behavioral profiles (e.g., Archibald et al., 2013; Astle et al., 2022; Conti-Ramsden et al., 1997). Notably, the latent clusters and profiles identified in these children's cognitive profiles often do not accord to the diagnostic labels they had been assigned. For example, in Astle et al.'s study, children with diagnoses

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

of autism, ADHD, dyslexia, dyscalculia, and DLD were represented in every cluster and there was little alignment of diagnostic classification with the data-driven groupings.

Debates about discrete categories, dimensions of variability, and heterogeneity have been rehearsed in more focused domains than whole cognitive profiles, for example, within reading development. In reading, profiles are identified according to the ability to read words with regular grapheme-phoneme correspondences, the ability to read words that are exceptions to these correspondences, and to read novel pseudowords. Languages with transparent orthographies do not have exception words, and here profiles of reading are distinguished by differences in reading accuracy versus reading rate.

In the reading domain, some researchers have argued that sub-types of dyslexics could be identified with particular problems reading exception words ('surface' dyslexics) or with reading novel pseudowords ('phonological' dyslexics (e.g., Castles & Coltheart, 1993). For example, Castle and Coltheart's regression-outlier method diagnosed pure subtypes by showing performance below age level in only exception word reading or only pseudoword reading, as well as mixed patterns. Different numbers of subtypes have been claimed (e.g., Chalmpe & Vlachos, 2025, claim three subtypes), while separate subtypes have been identified according to rate versus accuracy (Jabbour-Danial et al., 2024). Consistent with the dimensional approach, other researchers have claimed that subtyping is undermined by heterogeneous profiles (Zoubrinetzky et al., 2014).

Variability within narrower domains brings the phenomenon of heterogeneity within the scope of investigation via computational methods, since there are more likely to be established developmental models. This is the case with reading, where researchers have sought to understand how alterations to well-established models of typical development might capture subtypes or heterogeneity (e.g., Peterson et al., 2013; see Thomas & Karmiloff-Smith, 2003, for similar work in the domain of inflectional morphology). Here, developmental problems in specific mechanistic pathways or representations have been proposed to explain categorical subtypes of dyslexia. For example, disruptions to lexical-semantic information might impair the learning of exception words, while disruption to phonological information, or the pathway linking orthographic to phonological information, might impair the ability to generalize reading knowledge to novel forms (Harm & Seidenberg, 1999, 2004).

To date, this work has been limited in several ways. First, variability is usually captured by creating a model of (average) typical development on the one hand, and one or more parameter modifications to create a particular neurodevelopmental condition on the other. However, especially with the dimensional approach, empirical data address large populations, with mechanistic variations assumed to occur on continua within them. Second, to the extent that conditions such as dyslexia or DLD have genetic components which impact on neurocomputational properties, behavioral genetic studies now suggest there are polygenic causes of differences, with many DNA variants each contributing a small influence on the underlying substrate across diverse properties of neurocomputation (e.g., for reading, Doust et al., 2022; for speech and language, Benchek et al. 2021). Third differences in environmental cognitive stimulation (such as that sometimes linked to differences in socioeconomic status [SES]) may contribute to the variability observed in developmental outcomes, perhaps exaggerating, perhaps ameliorating genetic influences (e.g., Asadi et al., 2023). This source of variation also needs to be captured in models. Lastly, as a consequence of neglecting population-level accounts, domain models of variability have not explored whether latent profiles can be recovered by statistical approaches, as used in the dimensional approach, and therefore how latent profiles might be linked to underlying mechanistic causes of variation.

In this paper, we present a population-level model of the development of inflectional morphology, focusing as with reading on possible uneven profiles in the development of regular inflections, exception inflections, and generalization of regularities to novel forms. This model has previously been applied to exploring language delay, precocious language development, and the influence of SES on developmental trajectories (Thomas & Knowland, 2014; Thomas, Forrester & Ronald, 2013; Thomas, 2018). The model implements both environmental and genetic influences on development, the latter through polygenic influences on neurocomputation (Thomas, Forrester & Ronald, 2015).

Our goals were to: (1) use statistical techniques to assess whether latent profiles could be recovered from continuous dimensions of mechanistic variation; (2) compare these to subtypes defined via Coltheart and Castle's outlier method; (3) test the stability of latent profiles across development; (4) create a population with increased risk of genetic developmental atypicalities and establish how the profiles recovered by latent profile analysis were altered; (5) replicate this population to show how robust the recovered profiles were to stochastic factors in development.

Method

The model comprised populations of learners simulated using shallow artificial neural networks (ANNs). The model was trained on a domain drawn from language development, inflectional morphology (past tense). The model learned the mapping from phonological and lexical-semantic representations of verbs to phonological representations of their past tense (see [BLINDED] for details).

Training set: Models were trained on a set of 508 monosyllabic verbs, constructed using consonant-vowel templates and the phoneme set of English, with 557 input features and 62 output features. There were 410 regular verbs that formed their past tense by adding +ed, and 98 exception verbs split into three types, 20 no-change (nc) exceptions (e.g., hit-hit), 68 internal vowel-change (vc) exceptions (e.g., hide-hid), and 10 arbitrary (ar) exceptions (e.g., go-went).

Generalization was tested with 410 novel verb stems sharing two phonemes with existing regular verbs.

Model architecture: Shallow backpropagation-trained ANNs were used, either with two layers, three layers, or fully connected architectures, shown in Figure 1.

Hyper-parameter variability: In line with a polygenic understanding of the origin of cognitive variability, variations were implemented over 16 neurocomputational parameters: *Network construction*: architecture, number of hidden units, range of initial connection weight randomization, and sparseness of initial connectivity between layers. *Network activation*: unit threshold function, processing noise, and response accuracy threshold. *Network adaptation*: backpropagation error metric, learning rate, and momentum. As well as an overall learning rate, there were separate parameters modifying the learning rate between the lexical-semantic input units and the hidden units, and the phonological input units and the hidden units, potentially altering the relative balance of these sources of information during learning and therefore allowing more lexical or phonological strategies to past-tense acquisition (Joanisse & Seidenberg, 2003). *Network maintenance*: weight decay, pruning onset, pruning probability, and pruning threshold. Neurocomputational parameters were generated via an artificial genome that produced a probability distribution of each parameter across the population (see [BLINDED for details]). Distributions are shown in Figure 6. Each population comprised 1000 individuals. Each network was trained for 1000 epochs (presentations of the full training set).

Training set variability: Each network simulated a child raised in a given family, and families were assumed to vary in the richness of the language used. A training set was created for the past-tense information available in each family environment using a probabilistic filter that selected between 60% and 100% of the full training set.

Low-risk population: Figure 6 shows the probability distribution for the hyper-parameters for the base population.

High-risk population: Piloting revealed individual parameter manipulations that could produce specific deficits in rule generalization and in exception verb acquisition, as well as combinations of these deficits (Figure 2). Based on this piloting, changes were made to the population probabilities of hidden units (HU), sigmoid temperature in the unit activation function (TMP), general learning rate (GLR), phonological pathway learning rate (PLR), and lexical-semantic pathway learning rate (SLR) to reduces their values (see Figure 6). Using these revised probabilities, two high-risk populations were stochastically generated, each likely to show a greater proportion of developmental deficits.

Latent Profile Analysis (LPA): LPA was performed using the tidyLPA package (Rosenberg et al., 2019) in R (R Core Team, 2023). LPA is a type of latent variable model that is used to identify possible subgroups of participants to better understand sample heterogeneity (Sterba, 2013). Performance across five verb types at three time points (30, 100 and 1000 epochs) was analyzed. EEE models were fitted to performances at each time point to identify subgroups of individuals with similar performance levels. In an EEE model, the variances and covariances are estimated to be equal and can vary within the class but cannot vary between classes. Following Henson et al. (2007), the Sample-size-adjusted Bayesian Information Criterion was selected to determine the best fit number of subgroups.

Sub-grouping by outliers: Following Coltheart and Castles (1993), an alternative method was employed to determine subgroups. The low-risk population was used to derive the mean and standard deviation across regular verbs, rule generalization, and exception verb performance. Individuals who scored more than 1.5 standard deviations below each mean were classified as showing a deficit in that behavior. The same thresholds were then applied to the high-risk populations.

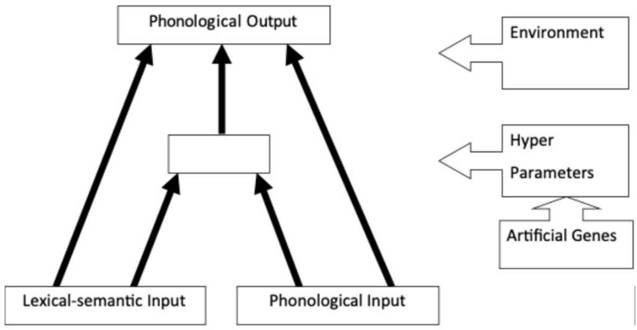

Figure 1: Model architecture

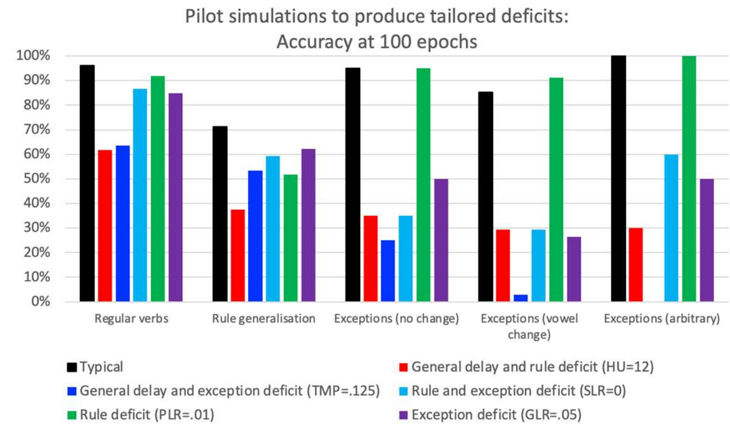

Figure 2: Individual parameter manipulations to identify ways to produce specific developmental deficits. Typical parameter values were TMP=1.0, HU=25, GLR=.125, PLR=.5, SLR=.5.

Results

The learning ability of each simulated individual was determined by the interaction of 16 dimensional parameters, while learning outcomes were also influenced by the richness of the environment to which the individual was exposed. For the low-risk population, LPA recovered several profiles across performance on regular verbs, rule generalization, and each exception verb type (Figure 3). Early in development, 12 profiles were distinguished by the algorithm; in mid development, this reduced to 8, and late in development this reduced to 5. The population therefore showed greater diversity in early profiles than in outcomes.

Early in development the outlier method, which defined deficits according to a cut-off, showed that individuals could be identified with specific deficits, either in exception verb performance or generalization performance, more rarely for regular verbs. However, general delay was a more common pattern than specific deficits (Figure 4, numbers in blue)

High-risk populations were created based on piloting to increase the frequency of uneven profiles and specific deficits. LPA results are shown in Figure 5. The manipulation was successful in lowering the overall performance of the population. For high-risk population #1, 4 latent profiles were identified early in development, 17 in mid-development, 13 in late development. In contrast to the low-risk population, early profiles were less differentiated, and diversity increased across development, particularly in the middle phase.

High-risk population #2 was generated stochastically using the same probability distribution of hyper-parameters, and new random association of networks to sampled training sets. This population showed a similar lowered level of overall performance. LPA recovered 2 latent profiles early in development, 9 in mid development, and 12 late in development. It exhibited a similar overall pattern of fewer latent classes early in development and then increasing diversity, but the detailed classes were different. For example, at the late time point, there was a unique class in population #1 showing a strong generalization deficit (class 10 in turquoise), while in population #2 there was a unique class showing a strong vowel-change exception deficit (class 1 in red).

Finally, the outlier approach demonstrated an inflation in the number of specific deficits and deficit combinations, with the deficit type inflating at different rates (Figure 4, numbers in red). There was little increase in selective generalization deficits, larger increases in regular verb deficits either in isolation or in combination with rule or exception, and a substantial increase in general deficits.

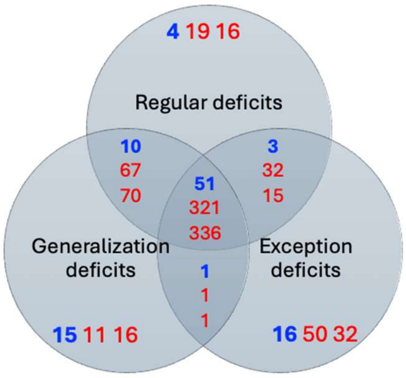

Figure 4. Patterns of deficit identified using the outlier method early in development (performance < 1.5 standard deviations below the mean for the low-risk population). Blue: low-risk; red: high-risk.

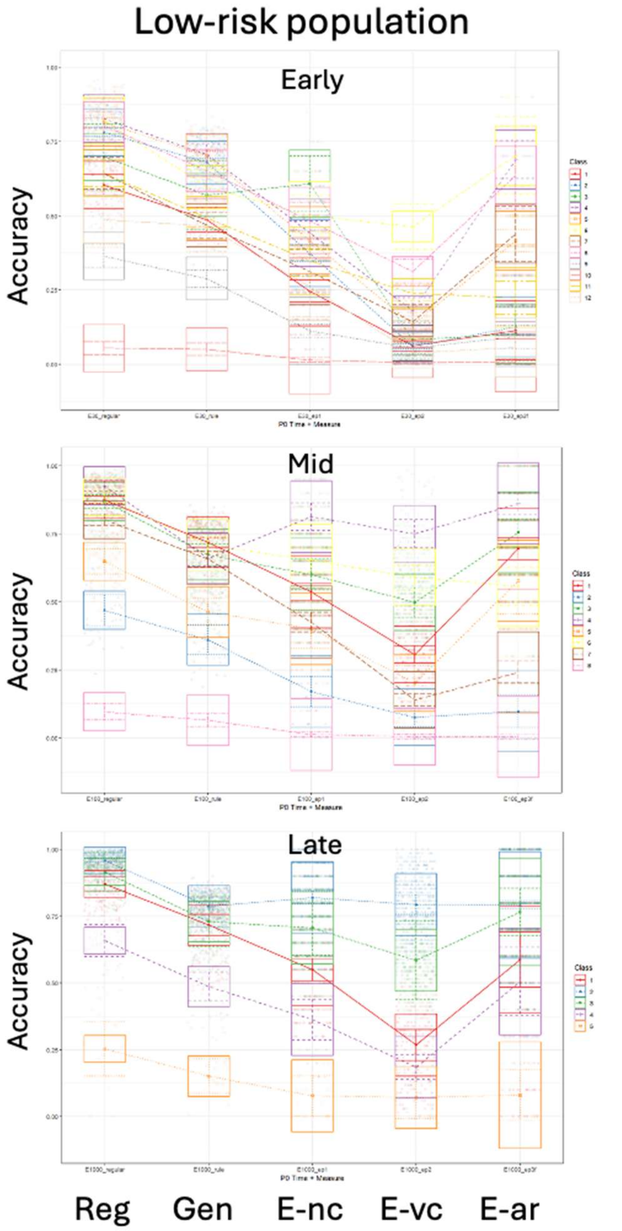

Figure 3. Profiles recovered by LPA at early, mid, and late points in development, across regular verb (Reg), novel verb generalization (Gen) and three exception verb types (E-)

Discussion

The history of computational modelling of deficits in neurodevelopmental conditions has lent itself more to understanding the mechanistic basis of categorical disorders. Researchers created a typical model, then changed some parameters to create an atypical model, implicitly assuming discrete differences and homogeneity amongst individuals with a condition. However, theory has now shifted to view neurodivergence as dimensional, with heterogeneity and comorbidity the rule rather than the exception.

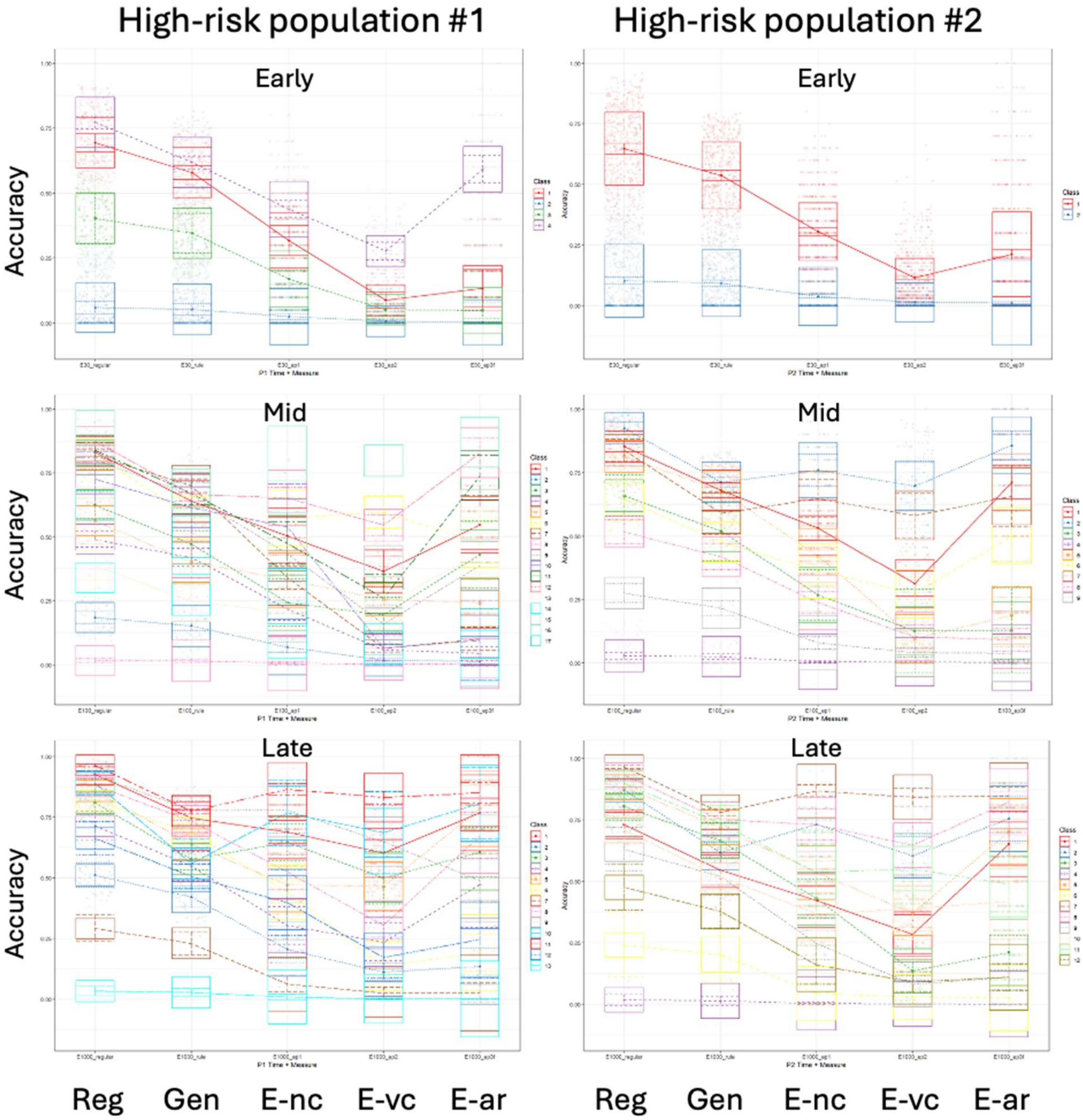

Figure 5. Profiles recovered by LPA for the high-risk populations across regular verb (Reg), novel verb generalization (Gen) and three exception verb types (E-)

The current work presents a mechanistic simulation framework that fits with the dimensional view, with underlying causes implemented both as polygenic, with many small influences on multiple neurocomputational parameters, and as partly a result of variable environments.

In a single domain with several distinguishable behaviors, we demonstrated that statistical techniques could be used to recover subgroups with different latent profiles. While the number of profiles might depend on the goodness-of-fit criterion used in a given statistical technique, we showed that with the same criterion, developmental change altered the number of profiles.

In a population at low risk of development deficits, there were more subgroups early in development and fewer later in development. Using outlier cut-off methods derived from neuropsychology, some small number of specific deficits could be observed, as has been reported in research on reading, but larger proportions showed combined deficits (Castles & Coltheart, 1993; Zoubrinetzky et al., 2014).

When we altered the hyper-parameters of the population to increase the risk of developmental deficits, we observed a shift in the latent profiles recovered by the statistical technique, demonstrating mechanistically the link between dimensional variations and developmental subtypes. In parallel, the outlier method showed increased numbers of specific deficits. We observed an influence of both pathway-specific parameters, such as those influencing lexical versus phonological strategies, and architecture-general parameters, such as global learning rate and hidden unit number, in producing uneven developmental profiles. This reconciles Harm and Seidenberg's model-based claim in 1999 that surface dyslexia (a greater deficit reading exception words) might represent a developmental delay in the reading system, and their later 2004 claim that it might instead constitute atypical division of labor between phonological and semantic routes for reading.

In line with Gillberg's (2010) ESSENCE framework, we saw atypicality compress and simplify early developmental trajectories in the high-risk populations, such that there were fewer latent profiles early in development, with increasing diversity as atypical constraints drew the trajectories apart – the reverse of the pattern seen in the low-risk population. Lastly, we looked at stochasticity. When the high-risk population was instantiated twice with the same probability distribution of parameters but in different individuals and with different random associations to variable environments, the broad patterns of latent profiles and outlier-identified specific deficits were similar, but the details differed.

The current modelling work comes with necessary caveats about the levels of simplification involved. It employed, for example, shallow ANNs with an artificial training set rather than deep neural networks exposed to naturalistic corpuses. The smaller scale was deployed to facilitate population-level approaches (here, N=3000 networks). And profiles were identified within the narrow domain of inflectional morphology, rather than the complex cognitive profiles considered in recent transdiagnostic studies.

Nevertheless, the model represents a shift towards building a mechanistic understanding of the origins of neurodevelopmental conditions that reflects both the new insights into the genetic and environmental influences on the emergence of uneven cognitive profiles, and the reality of those profiles in children as experienced by parents, educators, and clinicians.

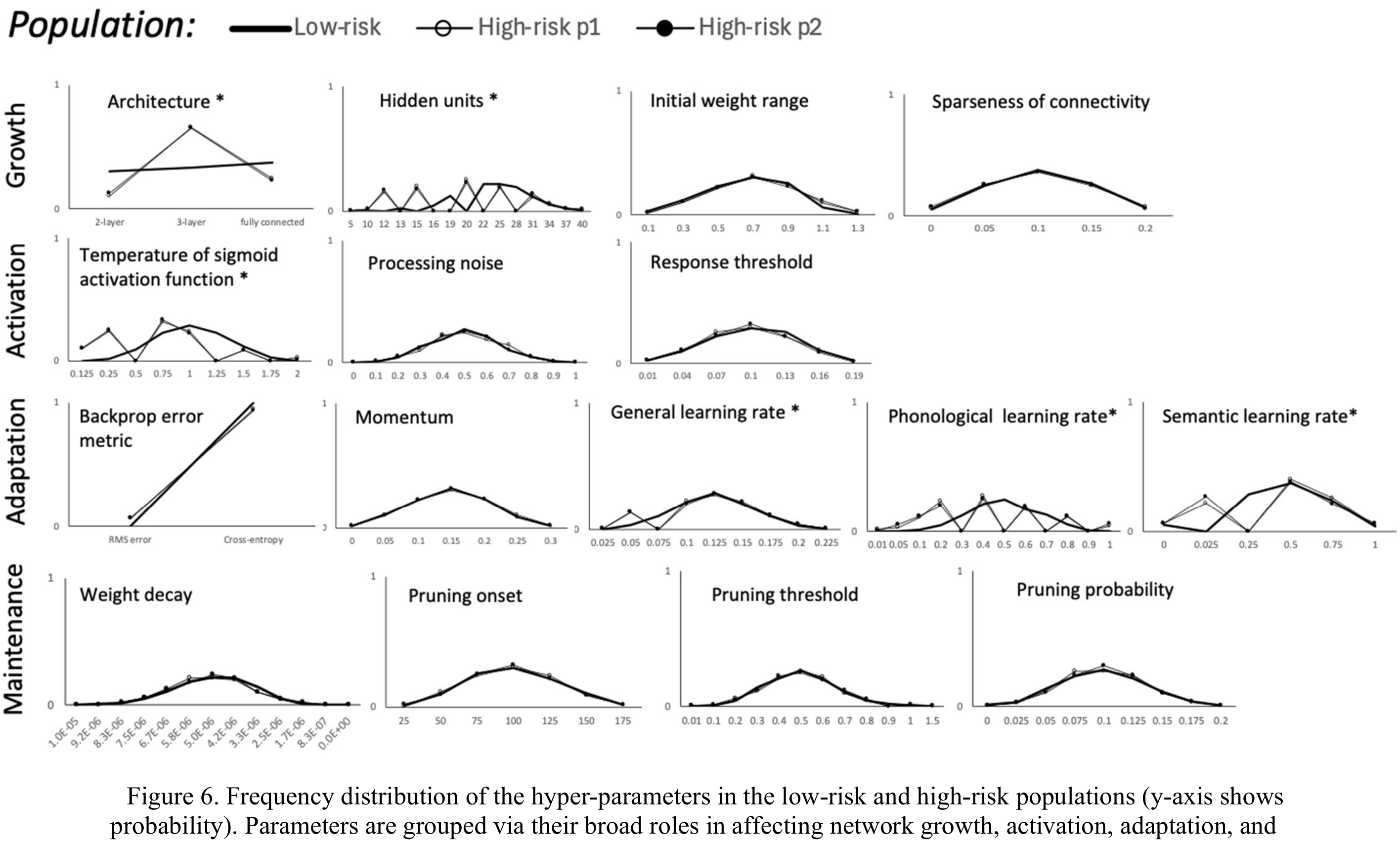

Figure 6. Frequency distribution of the hyper-parameters in the low-risk and high-risk populations (y-axis shows probability). Parameters are grouped via their broad roles in affecting network growth, activation, adaptation, and maintenance. Parameters with * are those where the distribution was altered for the high-risk population to increase the probability of uneven profiles, based on the piloting results shown in Figure 2. p1= population #1; p2 = population #2.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. https://doi.org/10.1176/appi.books.9780890425596

Archibald, L. M., Oram Cardy, J., Joanisse, M. F., & Ansari, D. (2013). Language, reading, and math learning profiles in an epidemiological sample of school age children. *PLoS One*, *8*, e77463.

Asadi, I. A., Kasperski, R., & Sarid, M. (2023). The cumulative effect of socioeconomic status and dyslexia on linguistic, cognitive and reading skills among Arabic-speaking children. *Dyslexia*, *29(2)*, 78-96. doi: 10.1002/dys.1735.

Astle, D. E., Holmes, J., Kievit, R., & Gathercole, S. E. (2022). Annual Research Review: The transdiagnostic revolution in neurodevelopmental disorders. *J Child Psychol Psychiatry*, *63(4)*, 397-417. doi: 10.1111/jcpp.13481.

Benchek, P., Igo, R. P., Voss-Hoynes, H. et al. (2021). Association between genes regulating neural pathways for quantitative traits of speech and language disorders. *npj Genom. Med.*, *6*, 64, https://doi.org/10.1038/s41525-021-00225-5

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, *47(2)*, 149–180.

Chalmpe, M. & Vlachos, F. (2025). Are there distinct subtypes of developmental dyslexia? *Front. Behav. Neurosci.*, *18*, 1512892. doi: 10.3389/fnbeh.2024.1512892

Conti-Ramsden, G., Crutchley, A., & Botting, N. (1997). The extent to which psychometric tests differentiate subgroups of children with SLI. *Journal of Speech, Language, and Hearing Research*, *40*, 765–777.

Doust, C., Fontanillas, P., Eising, E. et al. (2022). Discovery of 42 genome-wide significant loci associated with dyslexia. *Nat. Genet.*, *54*, 1621–1629. https://doi.org/10.1038/s41588-022-01192-y

Gillberg, C. (2010). The ESSENCE in child psychiatry: Early Symptomatic Syndromes Eliciting Neurodevelopmental Clinical Examinations. *Res Dev Disabil.*, *31(6)*, 1543-51. doi: 10.1016/j.ridd.2010.06.002.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106(3)*, 491–528.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, 111(3), 662–720.

Henson J. M., Reise S. P., & Kim K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling*, *14*, 202-226. https://doi.org/10.1080/10705510709336744

Hours, C., Recasens, C., & Baleyte, J-M. (2022). ASD and ADHD Comorbidity: What Are We Talking About? *Front. Psychiatry*, *13*, 837424. doi: 10.3389/fpsyt.2022.837424

Jabbour-Danial, M., Share, D. L., & Shalhoub-Awwad, Y. (2024). Not just phonology: a longitudinal study of dyslexia subtypes based on the distinction between reading accuracy and reading rate. *Front. Lang. Sci.*, *3*, 1390391. doi: 10.3389/flang.2024.1390391

Jackson, J. H. (1958). *Selected writings of John Hughlings Jackson (Vol. 1)*. In: Taylor J (ed). Staples Press, London.

Joanisse, M. F., & Seidenberg, M. S. (2003). Phonology and syntax in specific language impairment: Evidence from a connectionist model. *Brain and Language*, *86*(1), 40–56. https://doi.org/10.1016/S0093-934X(02)00533-3

Peterson, R. L., Pennington, B. F., & Olson, R. K. (2013). Subtypes of developmental dyslexia: testing the predictions of the dual-route and connectionist frameworks. *Cognition*, *126(1)*, 20-38. doi: 10.1016/j.cognition.2012.08.007.

R Core Team (2023). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/

Rosenberg, J. M., Beymer, P. N., Anderson, D. J., Van Lissa, C. J., & Schmidt, J. A. (2019). tidyLPA: An R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *Journal of Open Source Software*, *3*(30), 978.

Sterba, S. K. (2013). Understanding linkages among mixture models. *Multivariate Behavioral Research*, *48*(6), 775-815.

Tein, J. Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural equation modeling: a multidisciplinary journal*, *20*(4), 640-657.

Thomas, M. S. C. (2018). A neurocomputational model of developmental trajectories of gifted children under a polygenic model: When are gifted children held back by poor environments? *Intelligence*, *69*, 200-212. https://doi.org/10.1016/j.intell.2018.06.008

Thomas, M. S. C., Forrester, N. A., & Ronald, A. (2015). Multi-scale modeling of gene-behavior associations in an artificial neural network model of cognitive development. *Cognitive Science*, *40(1)*, 51-99. doi: 10.1111/cogs.12230

Thomas, M. S. C. & Karmiloff-Smith, A. (2003). Modelling language acquisition in atypical phenotypes. *Psychological Review*, *110(4)*, 647-682. https://doi.org/10.1037/0033-295X.110.4.647

Thomas, M. S. C. & Knowland, V. C. P. (2014). Modelling mechanisms of persisting and resolving delay in language development. *Journal of Speech, Language, and Hearing Research*, *57(2)*, 467-483. doi:10.1044/2013_JSLHR-L-12-0254.

Thomas, M. S. C., Ronald, A., & Forrester, N. A. (2013). Modeling socio-economic status effects on language development. *Developmental Psychology*, *49(12)*, 2325-43. doi:10.1037/a0032301.

Zoubrinetzky, R., Bielle, F., & Valdois, S. (2014). New Insights on Developmental Dyslexia Subtypes: Heterogeneity of Mixed Reading Profiles. *PLoS ONE*, *9(6)*, e99337. doi:10.1371/journal.pone.0099337