UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

## Title
Modeling Face Recognition Challenges in Autism Spectrum Disorder: A CNN-Based Approach

## Permalink
https://escholarship.org/uc/item/8sr2592s

## Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

## Authors
Wang, Xijing
Chen, Lang

## Publication Date
2025

## Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

# Modeling Face Recognition Challenges in Autism Spectrum Disorder: A CNN-Based Approach

**Xijing Wang (xwang20@scu.edu)**
Department of Mathematics and Computer Science, 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**

Computational modeling has been a crucial tool in cognitive science to understand human cognitive functions and impairments in neurocognitive disorders. Convolutional Neural Networks (CNNs) exhibit striking similarities to human visual processing systems for object recognition, making them a powerful tool for studying visual processes. In this study, we examined the neurobiological theories, namely, the Excitation/Inhibition (E/I) Imbalance and Internal Noise (IN) in explaining face recognition challenges in autism spectrum disorder (ASD) using CNNs, and revealed that over-excitation and increased noises in the CNNs led to compromised performance on face recognition and atypical patterns of internal representations of face stimuli. This approach enables systematic comparisons between typical and atypical cognition, offering a theory-driven perspective to investigate cognitive challenges and their neurocognitive mechanisms with a computational approach.

**Keywords:** Computational modeling; Face recognition; Autism spectrum disorders; Neural representations; Excitation/Inhibition imbalance; Internal noise

## Introduction

Computational models, especially deep neural networks, have been a cornerstone in cognitive science, providing a structured framework for understanding human cognition and brain functions (Chen et al., 2017; Farah & McClelland, 1991; Kriegeskorte & Douglas, 2018; McClelland & Rogers, 2003). Recent endeavors showed that computational models cannot only simulate cognitive behavior and deficits (Dilkina et al., 2008; Seidenberg, 2005) but also provide a window to examine individual variabilities in internal neural representations (Chen et al., 2017; Mehrer et al., 2020). Specifically, the Convolutional Neural Network (CNN) model is a dominant tool to approximate visual processing and object recognition with reasonable resemblance to the ventral visual pathway in humans (Hubel & Wiesel, 1962), with populational activations(LeCun et al., 2015), and processing hierarchy (Felleman & Van Essen, 1991). Studies have shown that CNN models can successfully explain the variance in neural responses in the inferotemporal cortex neurons for image recognition (Cadieu et al., 2014) and show high correlations with brain responses in the visual cortex of humans as well (Yamins et al., 2014). In the current study, we will use a theory-driven approach to simulate face recognition in CNN models and manipulate key model parameters based on the Excitation/Inhibition (E/I) imbalance and internal noise (IN) theories. By modulating these biologically meaningful parameters, we would like to examine the predictions of E/I imbalance and IN theories to explain the face recognition challenges in autism spectrum disorder (ASD).

ASD is characterized by socio-communicative challenges (CDC - Centers for Disease Control and Prevention et al., 2018), but autistic individuals also show cognitive challenges in face recognition (Minio-Paluello et al., 2020; Weigelt et al., 2012), a crucial factor for social interaction (Klin et al., 2002). Neuroimaging studies have also shown atypical neural activations, representations, and functional connectivity in autistic individuals during face processing (Lynn et al., 2018; O'Hearn et al., 2020; Pierce, 2001; Schultz, 2005). However, behavioral and neuroimaging studies in human subjects can hardly reveal the neurocognitive mechanism of atypical face recognition in ASD (Jiang et al., 2022) since these studies could not manipulate neurobiological factors in humans and failed to provide causal implications.

For example, the excitation/inhibition imbalance (E/I imbalance) theory hypothesizes that the imbalanced ratio between excitatory (e.g., glutamatergic) and inhibitory (e.g., GABAergic) neural signals underlie many developmental disorders, including ASD (Rubenstein et al., 2010). Another theory, known as internal noise (IN) theory (Park et al., 2017), provides a different lens of the cognitive challenges observed in ASD. Specifically, the increased internal noise leads to unusually large fluctuations in neural responses, resulting in unreliable and less predictable representations of the environment, resulting in neurocognitive deficits in ASD (Dinstein et al., 2015).

These two theories also have explicit predictions on the face recognition challenges in autistic individuals. According to the E/I imbalance, excessive excitatory signals could lead to diffuse activations in the brain, resulting in undifferentiated neural representations (Chen et al., 2021) for faces and reduced ability for face recognition. Overexcitation will likely result in increased neural representation similarity for faces of different persons (between-individual similarity)



In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

but have little impact on the neural representation similarity for faces of the same person (within-individual similarity). Meanwhile, increased internal noises in the brain will create unstable neural responses with random fluctuations.

both between-individual and within-individual similarity of neural representations should be reduced, hampering the ability to recognize faces. Thus, although not mutually exclusive, the two theories have different predictions of underlying neural representations that result in face recognition challenges in ASD. The CNN models provided the ideal vehicle for us to test the predictions of E/I imbalance and IN theories on the behaviors and internal representations of faces in a well-controlled learning system, potentially shedding light on the neurocognitive mechanisms of face recognition challenges in autistic individuals.

# Method

## General Model Structure

We implemented standardized convolutional neural network (CNN) models (Farfade et al., 2015; Krizhevsky et al., 2012) for face processing with one input layer, three convolutional blocks, a flattening layer, and a dense layer (i.e., the output layer). Each of the convolutional blocks consisted of a convolutional, normalization, pooling (size = 2*2), and dropout (rate = 0.2) layers, primarily serving the purpose of hierarchical feature extraction (**Figure 1**). The convolutional layer applies matrix operations to detect features of inputs (images) such as edges, corners, and curves. The normalization and max pooling layers stabilize the activations and retain the most salient features by reducing spatial dimensions. The dropout layer serves as a regularization technique by randomly deactivating 20% of neurons during training, thereby encouraging the model to learn more robust and generalizable feature representations rather than overfitting to specific neuron activations. After the three convolutional blocks, the multi-dimensions feature maps were flattened into a one-dimensional vector to enable classification. In the last dense layer, the flattened output was processed by the SoftMax activation function to produce a probability value for the identity of each person (10 units, each for one individual's identity). The filter size of each convolutional layer was set to 16, 32, and 64, as practiced in the literature. When we tested the hypotheses of the E/I imbalance and IN theories in the CNN models, we manipulated two parameters: the slope of the Modified ReLU function of convolutional layer in each convolutional block (**F1**) and the standard deviation of the Gaussian noises (**F2** and **F3**) added to all the weights (after each convolutional block), respectively. We will explain the details in the training and testing procedure section.

$$Modified\ ReLU: f(x) = max(0, \alpha x) \quad \textbf{(F1)}$$

$$Y_{noise} = Y_{original} + N(0, \sigma^2) \quad \textbf{(F2)}$$

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{(x-\mu)^2}{2\sigma^2}) => N(0, \sigma^2) \quad \textbf{(F3)}$$

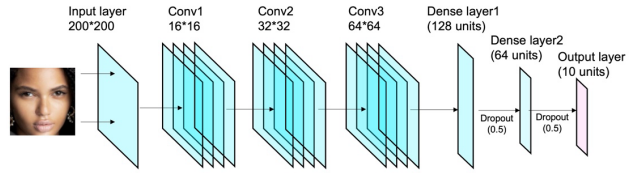


**Figure 1**. The architecture of the CNN models.

This setup allows us to analyze how neurobiologically meaningful parameters in the CNN models may affect learning and representations of faces and will enable us to compare the behaviors of models with findings from the human participants.

## Training and testing sets

The training and testing set of face stimuli comprised 50 selected images from 10 celebrities, with five images representing each celebrity. All images were gathered from the internet, including five males and five females, and from diverse demographic backgrounds such as age, race, and ethnicity. Only images depicting a full face from a front perspective were selected, and the face part was cropped for training and testing. All face images were cropped and standardized to include only face features with a dimension of 200-by-200 pixels from the original images.

## Training and Testing Procedures

The standard training procedures included a face recognition task, in which the CNN models were presented with face stimuli (4 out of 5 from each individual) and trained to classify the individual identities at the output layer in a localist fashion. One image from each individual was randomly reserved for testing. All models were trained with the Adam optimizer (Kingma & Ba, 2015) with a learning rate of 0.001 and binary cross-entropy loss function. The output layer utilized the softmax function to generate the final label according to each input stimulus pattern (i.e., each image). All the CNN models were trained for 1,000 epochs and 20 independent loops with random initial weights were run to resemble individual differences under each experimental condition.

Since we were mostly interested in how varying model parameters could lead to different learning performance and internal representations, we focused our analysis on trained items (i.e., four trained images of each individual). For each loop under each condition, we scored the accuracy by dividing the total number of corrected predictions by the total number of predictions (i.e., 4 X 10 = 40 total predictions) at each training epoch. Then, we compared the mean accuracy of face recognition between different conditions.

**E/I imbalance simulation.** As mentioned, we manipulated the $\alpha$ value in the modified ReLU function (**F1**), a scaling factor that controls the positive output values of all units.

Conventionally, the $\alpha$ value was set to 1. Through pilot testing, we found that the CNN models with $\alpha$ = 0.05 represented the optimal learning outcomes for our dataset, so we chose it as a typical level of excitation (i.e., an optimal balance between excitatory and inhibitory signals). For the over-excitation condition, we set the value to 0.5, ten times larger than the optimal $\alpha$ value. We also included an under-excitation condition as a comparison, in which the $\alpha$ value was set to 0.005, one-tenth of the optimal value. Note that we did not aim to argue for a numerical equivalence of the over- and under-excitation conditions between the CNN model simulations and the biological conditions in humans. These values were chosen to merely demonstrate the impact of a single parameter that bears neurobiological implications to the E/I imbalance theory. We hypothesized that an over-excitation in CNN models would hinder the face recognition ability by confusing the internal representations of the faces of different individuals.

**IN simulation.** To test the predictions of the Internal Noise (IN) theory in CNN models, we introduced a Gaussian noise layer after each convolutional block, adding intrinsic but random noises into the signals after the dropout layers and before the next processing layers. In this way, the output of each unit from the dropout layers was offset by a value randomly sampled from a Gaussian distribution with a standard deviation of $\sigma$ (**F3**) and a mean of 0. For the typical condition, we set the $\sigma$ = 0 as in a system with minimal disturbance. Then, we changed $\sigma$ to 0.5 (a mildly noised condition) or 5 (a severely noised condition) to examine the influence of the additional internal noise on learning accuracy as well as the pattern separation in the internal representations. We hypothesized that increased noises would lead to noised representations of faces, resulting in reduced similarity of faces of the same individual, but increased similarity of faces of different individuals based on randomly activated units.

## Representational Similarity Analysis

Following the standard practice in relevant computational and neuroimaging research (Chen et al., 2021; Kriegeskorte et al., 2008; O'Hearn et al., 2020), we generated the similarity matrix by calculating the Pearson's *r* between every pair of trained images (4 images for each of the 10 individuals) over the units in the flattened layer. We chose the flattened layer because it was the last layer before the output layer, so it contained the highest level of the internal representations of all trained images in the CNN models. We then computed the mean correlation for images of the same individual (within-individual similarity) and the mean correlation for images of different individuals (between-individual similarity). Success in face recognition requires an ability to recognize the same person across images, i.e., a high within-individual similarity, and also the ability to differentiate images of different individuals, i.e., a low between-individual similarity. These measures could provide insights into how the CNN models generalize across images for the same individuals and different images of different individuals. Furthermore, comparing the within- and between-individual similarities across different conditions of E/I or IN conditions, we could assess the impact of E/I and IN on the ability to represent face stimuli within the CNN models.

After the models were trained, we tested all images and extracted the activation values of all units in the flattened layer. We then calculated the Pearson correlation between all possible pairs of the images based on the activation values of 64 units in the flattened layer. The similarity matrices of each loop after 500, 750, and 1,000 epochs were computed, and the mean within- and between-individual similarity was calculated separately. A high similarity value suggests overlapping representations whereas a low similarity value suggests separable representations in the CNN models.

# Result

## Accuracy of face recognition

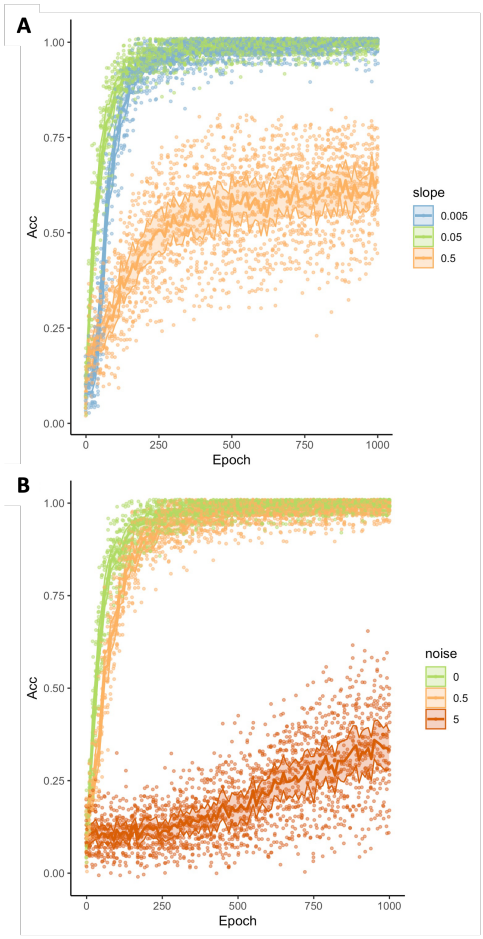


**Figure 2**. Accuracy of face recognition on trained images over 1,000 epochs for E/I imbalance simulations (**A**) and IN simulations (**B**). Individual points represent the accuracy of each loop, and the thick solid line represents the average for each condition. The shaded area around the thick lines represents the 95% confidence interval.

The results demonstrated that both Excitation/Inhibition (E/I) imbalance and Internal Noise (IN) significantly affected learning efficiency and recognition accuracy in CNNs, reflecting similar challenges in autistic individuals in face recognition. CNNs trained under over-excitation conditions (i.e., high ReLU slope) exhibited reduced recognition accuracy (**Figure 2A**), even after extended training of 500 epochs, $F(2, 57) = 225.25$, $p < 0.001$. In contrast, under-excitation (low ReLU slope) resulted in slower initial learning (epoch = 100, $p < 0.001$), but the model adapted over time and reached comparable accuracy by 300 epochs ($p = 0.905$).

Similarly, Internal Noise significantly disrupted learning stability, particularly under high-noise conditions (SD = 5), which consistently impaired face recognition in the CNN models (at 500 epochs, $p < 0.001$; **Figure 2B**). Mild noise (SD = 0.5) caused learning inefficiencies early in training (at the epoch of 100, $p < 0.001$), but its effects diminished over time ($p = 0.315$ at 300 epochs). This suggests that while moderate neural variability may be tolerable or even adaptive, excessive noise prevented the formation of stable representations of faces and impaired the ability in face recognition, mirroring the behaviors of autistic individuals.

Overall, these findings validate CNNs as a computational model for simulating neurobiological mechanisms on the disrupted learning of human faces. Over-excitation and high internal noise hampered cognitive functions in general, while under-excitation and mild noise levels in the learning system can be overcome over time.

## Neural Representation Similarity

Next, we examined the neural representation similarity (NRS) between images, and we especially investigated the NRS of images of the same individual identity (within-individual) vs. the NRS of images from different individual identities (between-individual). Although the learning accuracy was stable after 500 epochs of training for both simulations, it was surprising to observe that within- vs. between-individual NRS were not quite separate from each other for models with optimal settings (i.e., slope = 0.05 and Gaussian noise = 0). Therefore, we examined the change of NRS from 500, 750, and 1,000 training epochs (Results from 1,000 training epochs in **Figure 3**).

The within-individual NRS values were higher than between-individual NRS values across different training epochs, but the separation became larger with more training. When the model was over-excited (i.e., slope = 0.5), the between-individual NRS seemed to be high even after the model was trained for 1,000 epochs. A mixed ANOVA with Slope (0.005, 0.05, and 0.5) as a between-subject factor and with NRS Type (Within vs. Between) as a within-subject factor revealed significant main effects of both Slope and Type for all three training epochs. At 500 epochs, the optimal slope (i.e., 0.05) showed the highest within-individual NRS compared to the other two conditions (both $p$s < 0.05, HSD adjusted), but also the highest between-individual NRS (both $p$s < 0.01, HSD adjusted). In contrast, after 750 epochs of training, as well as seen for 1,000 epochs of training, the over-excitation models demonstrated the lowest within-individual NRS but the highest between-individual NRS compared to the optimal and under-excitation conditions (**Table 1**). These findings suggested a weak ability to represent images of the same individual similarly but also a difficulty in differentiating images of different individuals at the representational level.

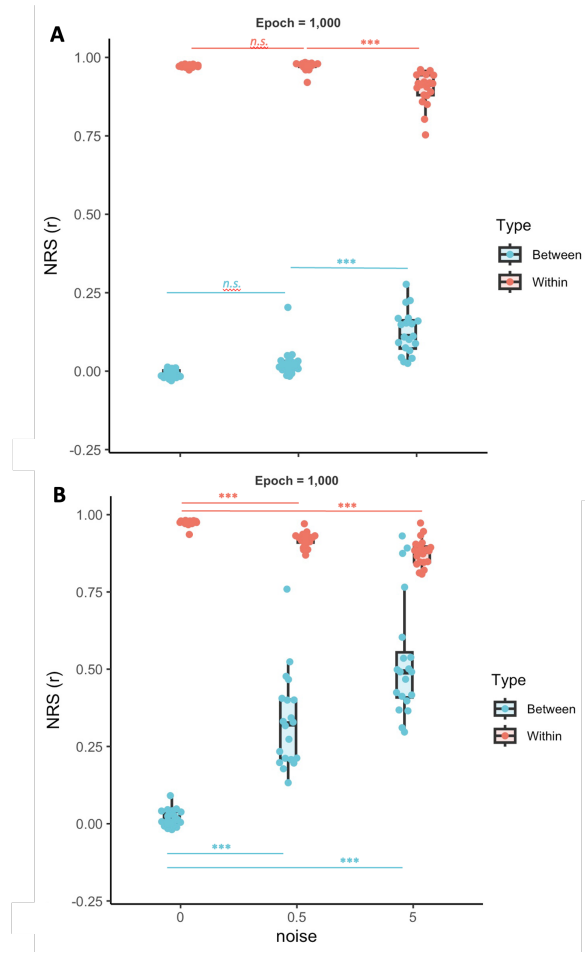


**Figure 3**. Boxplot of the neural representational similarity (NRS) for between-individual (Blue) and within-individual (Red) for E/I imbalance simulations (**A**) and IN simulations (**B**) at 1,000 training epochs.

The results for NRS with IN simulation showed a similar pattern (**Figure 3B**): both within- and between-individual NRS values were high after 500 epochs of training, and the separation was evident with more training. At 750 epochs, the optimal condition (noise SD = 0) showed the highest within-individual NRS with the lowest between-individual NRS, whereas the results for models with some noise (SD = 0.5 or 5) still demonstrated relatively high levels of between-individual NRS. At 1,000 epochs, there was a graded effect

that the larger the noises applied to the weights, the less separation we observed for the within- and between-individual NRS values (**Table 1**). This finding suggests that increased internal noises in a learning system led to weak representations of face images in a monatomic way: Increased noise resulted in a combination of difficulties in representing faces of the same individual similarly as well as representing faces of different individuals differently.

**Table 1**. Mean NRS in E/I imbalance and IN simulations.

| Epoch | Type | Optimal (0.05) | Under- (0.005) | Over- (0.5) |
|---|---|---|---|---|
| 500 | Within | 1.00 (0.00) | 0.97 (0.02) | 0.83 (0.04) |
| | Between | 0.96 (0.03) | 0.84 (0.12) | 0.33 (0.13) |
| 750 | Within | 0.97 (0.01) | 0.96 (0.01) | 0.9 (0.03) |
| | Between | 0.07 (0.06) | 0.02 (0.02) | 0.14 (0.07) |
| 1000 | Within | 0.97 (0.01) | 0.97 (0.00) | 0.90 (0.05) |
| | Between | 0.03 (0.05) | -0.01 (0.01) | 0.12 (0.07) |
| **Epoch** | **Type** | **Optimal (0)** | **Mild (0.5)** | **High (5)** |
| 500 | Within | 1.00 (0.00) | 1.00 (0.00) | 0.92 (0.06) |
| | Between | 0.96 (0.02) | 0.99 (0.01) | 0.80 (0.15) |
| 750 | Within | 0.97 (0.01) | 0.94 (0.02) | 0.87 (0.04) |
| | Between | 0.07 (0.04) | 0.62 (0.15) | 0.59 (0.14) |
| 1000 | Within | 0.97 (0.01) | 0.92 (0.02) | 0.88 (0.04) |
| | Between | 0.02 (0.03) | 0.33 (0.15) | 0.53 (0.19) |

## Discussion

Our findings suggested that both E/I imbalance and internal noise were associated with reduced ability in face recognition in CNN models. Specifically, under-excitation led to inferior face recognition early during the learning, but over-excitation led to sustained deficits in face recognition. However, increased internal noises were associated with impaired face recognition in a graded way: the larger the noise was, the poorer the learning was. For internal representations, it was shown that over-excitation resulted in slightly reduced within-individual similarity and increased between-individual similarity, whereas under-excitation potentially reduced both within-individual and between-individual similarities, i.e., more differentiated neural representations. By contrast, the increased internal noise led to a monotonic change with decreased within-individual similarity but increased between-individual similarity, suggesting a generally undifferentiated neural representation of faces.

Our results based on E/I imbalance simulations were largely consistent with the predictions and existing literature. As predicted by the E/I imbalance theory, ASD is associated with excessive excitatory signals in the neural system, leading to cognitive challenges (Foss-Feig et al., 2017; Rubenstein et al., 2010; Yizhar et al., 2011). In our E/I imbalance simulations, we changed the slope of the ReLU function to approximate the excessive excitatory signals in the human brain (Krossa et al., 2023). We showed that the over-excitation (slope = 0.5) led to delayed learning of face identity and reached the plateau after 500 training epochs. This is consistent with the growing body of literature that autistic individuals have difficulties in face recognition (Dimitriou et al., 2015; Hauck et al., 1998; Liu, Chen, et al., 2023; Pallett et al., 2014; Stantić et al., 2022; Suri et al., 2021) and even memory in general (Bowler et al., 2011; Desaunay et al., 2020; Griffin et al., 2022; Lind & Bowler, 2010; Williams et al., 2006). Our simulations further suggested that over-excitation in the learning system resulted in atypical neural representations of faces compared to the optimal condition, particularly for the faces of different individuals, i.e., elevated levels of between-individual similarity, as well as a decrease in within-individual similarity, i.e., faces of the same individual. Previous studies have indeed suggested that neural representational similarity in the right fusiform face area (FFA) for the same type of stimuli (e.g., faces and cars) was lower in ASD than in non-ASD (O'Hearn et al., 2020), and autistic children showed reduced neural representation similarity after learning (Liu, Chang, et al., 2023). However, our results further suggest that the neural representational similarity of different kinds should also be higher. Therefore, the reduced face recognition ability due to over-excitation is likely a result of an inability to differentiate the faces of different individuals. These results are probably due to unnecessarily activated units in the different CNN layers, leading to using many overlapping units to represent the faces of different individuals (i.e., lack of sparse representations). This finding is also consistent with the recent study suggesting that a subgroup of autistic children showed atypical pattern separation memory for daily objects with more diffused memory representations for similar memory experiences (Chen et al., 2024).

Interestingly, the under-excitation condition (slope = 0.05) showed mild influence on the accuracy of face recognition only at the early stage of learning and even more differentiated representations of the faces of different individuals (at 500 epochs, optimal: 0.96 vs under-excitation: 0.84; $p < 0.01$). Thus, over- and under-excitation seemingly led to opposite patterns of neural representations, which is consistent with the recent findings that the pattern separation memory in ASD is heterogeneous, and there are subgroups with both over-differentiated and under-differentiated representations of memory experience (Chen et al., 2024). This pattern potentially represents a large subgroup of autistic individuals who have more isolated memory experiences and a lack of generalization ability across similar experiences (Dovgopoly & Mercado III, 2013; South et al., 2015). Thus, the predictions from E\I imbalance were largely supported by our simulations and consistent with the literature on face recognition in ASD. More importantly, our simulation based

on E/I imbalance could potentially explain the different profiles of atypical face recognition and memory representations in ASD.

The results from IN simulations were also consistent with the hypothesis that increased noises will lead to lower face recognition ability (Davis & Plaisted-Grant, 2015; Park et al., 2017; Vilidaite et al., 2017). However, instead of observing reduced representation similarities overall, we found that the increased noise weakened the representations of faces of the same individual, i.e., less similar, but also confused the representations of faces of different individuals (i.e., more diffused representations). This finding is largely similar to the E/I simulations with over-excitations and contradictory to our hypothesis. One possible explanation is that the random noises added to the learning system could lead to units in CNNs being activated randomly. Once these units were activated, they were utilized to process the faces of different individuals, leading to overlapping representations of the faces of different individuals on these randomly activated units. At the same time, the randomly activated units also led to less overlapping representations of the faces of different individuals since they were unnecessary to represent the faces of the same individual. Our IN and E/I simulation (over-excitation) seems to suggest the intrinsic link between the theories of E/I imbalance (Rubenstein et al., 2010) and internal noise (Park et al., 2017). The large internal noise in the neural system of ASD may be a result of excessive excitatory signals. If so, although these two theories may capture different features of the neurobiological dysfunctions of ASD, they may be two sides of the same coin. However, based on our simulation, the effect of the internal noise was rather monotonic, so weaker/more diffused representations were found with increased levels of noise. Unlike the results from E/I imbalance, it failed to explain the full spectrum of atypical memory abilities and representations observed with both over- and under-differentiated memory representations in ASD revealed in recent studies (Chen et al., 2024; South et al., 2015).

Using a computational approach, our studies suggested that E/I imbalance and increased internal noise in the neural system could underlie the face recognition challenges in autistic individuals. Using a theory-driven approach, we were able to demonstrate the utility and power of using computational models to understand human cognition and neurocognitive mechanisms of cognitive disorders(Farah & McClelland, 1991; Plaut et al., 1996; Rogers et al., 2004), as well as their ability to bridge the behavioral and neuroimaging studies to examine individual differences (Chen et al., 2017; Dilkina et al., 2010; Mehrer et al., 2020; Woollams et al., 2007). However, we would emphasize again that by no means do we argue for a neurobiological faithfulness of these simulations and parameters in CNN models. In fact, we only consider this theory-driven approach a tool to test predictions of theories in a well-controlled learning system that can hardly be done in human research. We need converging evidence from behavioral, neuroimaging, and computational approaches to establish, test, and revise key theories of human cognition and to understand the neurobiological mechanisms of cognitive challenges in clinical populations such as ASD. We believe that this integrated approach can be easily applied to other domains to advance our knowledge of other types of developmental and neurobiological disorders, as well as their cognitive challenges. CNNs are well-suited for modeling face recognition due to their spatial feature extraction capabilities. Future research could extend this framework to a broader range of biologically inspired neural architectures. For instance, spiking neural networks (SNNs) have shown promise in modeling decision-making processes and real-time neural dynamics due to their temporal coding and bio-plausible structure (Maass, 1997; Xu et al., 2023). Similarly, recurrent neural networks (RNNs) have been successfully applied to model working memory and sequential cognitive processes. For example, recent work has demonstrated that RNNs with transient trajectories can replicate the neural dynamics observed in human working memory tasks (Liu et al., 2024). These architectures could provide complementary insights into the broader neurocognitive mechanisms involved in ASD and support future explorations beyond face processing. Moreover, we also do not argue that these two neurobiological factors are the only reasons why face recognition is challenging in ASD. These theories may only explain certain subtypes of face recognition challenges in ASD. Other possibilities, such as lack of social motivation (Klin et al., 2002; Mundy & Stella, 2000) or dysfunctions in the rewarding system for social stimuli (Abrams et al., 2013, 2019), could still result in face recognition challenges in ASD with seemingly similar behavioral outcomes, but different underlying mechanisms. Future work should examine how these factors interact and jointly affect the different phenotypes of ASD and their heterogeneous presentations of cognitive and social challenges (Chen et al., 2019; Georgiades et al., 2013; Jacob et al., 2019; Lynch et al., 2013; Ring et al., 2008).

## References


Abrams, D. A., Lynch, C. J., Cheng, K. M., Phillips, J., Supekar, K., Ryali, S., Uddin, L. Q., & Menon, V. (2013). Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proceedings of the National Academy of Sciences*, *110*(29), 12060–12065. https://doi.org/10.1073/pnas.1302982110

Abrams, D. A., Padmanabhan, A., Chen, T., Odriozola, P., Baker, A. E., Kochalka, J., Phillips, J. M., & Menon, V. (2019). Impaired voice processing in reward and salience circuits predicts social communication in children with autism. *ELife*, *8*, e39906. https://doi.org/10.7554/eLife.39906

Bowler, D., Gaigg, S., & Lind, S. (2011). Memory in autism: Binding, self and brain. *Researching the Autism Spectrum: Contemporary Perspectives*, 316–346. https://doi.org/10.1017/CBO9780511973918.013

Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., & DiCarlo, J. J. (2014). Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. *PLoS Computational Biology*, *10*(12), e1003963–e1003963. https://doi.org/10.1371/journal.pcbi.1003963

CDC - Centers for Disease Control and Prevention, Prevention, C. D. C.-C. for D. C. and, Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Rosenberg, C. R., & White, T. (2018). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR. Surveillance Summaries*, *67*(6), 1–23. https://doi.org/10.15585/mmwr.ss6706a1

Chen, L., Abrams, D. A., Rosenberg-Lee, M., Iuculano, T., Wakeman, H. N., Prathap, S., Chen, T., & Menon, V. (2019). Quantitative Analysis of Heterogeneity in Academic Achievement of Children With Autism. *Clinical Psychological Science*, *7*(2), 362–380. https://doi.org/10.1177/2167702618809353

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., Liu, J., Kang, J. B., Rosenberg-Lee, M., Abrams, D. A., & Menon, V. (2024). Atypical pattern separation memory and its association with restricted interests and repetitive behaviors in autistic children. *Autism*, 13623613231223354. https://doi.org/10.1177/13623613231223354

Davis, G., & Plaisted-Grant, K. (2015). Low endogenous neural noise in autism. *Autism*, *19*(3), 351–362. https://doi.org/10.1177/1362361314552198

Desaunay, P., Briant, A. R., Bowler, D. M., Ring, M., Gérardin, P., Baleyte, J.-M., Guénolé, F., Eustache, F., Parienti, J.-J., & Guillery-Girard, B. (2020). Memory in autism spectrum disorder: A meta-analysis of experimental studies. *Psychological Bulletin*, *146*(5), 377.

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

Dimitriou, D., Leonard, H. C., Karmiloff-Smith, A., Johnson, M. H., & Thomas, M. S. C. C. (2015). Atypical development of configural face recognition in children with autism, Down syndrome and Williams syndrome. *Journal of Intellectual Disability Research*, *59*(5), 422–438. https://doi.org/10.1111/jir.12141

Dinstein, I., Heeger, D. J., & Behrmann, M. (2015). Neural variability: friend or foe? *Trends in Cognitive Sciences*, *19*(6), 322–328. https://doi.org/10.1016/j.tics.2015.04.005

Dovgopoly, A., & Mercado III, E. (2013). A connectionist model of category learning by individuals with high-functioning autism spectrum disorder. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(2), 371–389.

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

Farfade, S. S., Saberian, M., & Li, L.-J. (2015). Multi-view Face Detection Using Deep Convolutional Neural Networks. *Cornell University Library*, 643–650. http://arxiv.org/abs/1502.02766

Felleman, D. J., & Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1

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

Georgiades, S., Szatmari, P., Boyle, M., Hanna, S., Duku, E., Zwaigenbaum, L., Bryson, S., Fombonne, E., Volden, J., Mirenda, P., Smith, I., Roberts, W., Vaillancourt, T., Waddell, C., Bennett, T., & Thompson, A. (2013). Investigating phenotypic heterogeneity in children with autism spectrum disorder: A factor mixture modeling approach. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(2), 206–215. https://doi.org/10.1111/j.1469-7610.2012.02588.x

Griffin, J. W., Bauer, R., & Gavett, B. E. (2022). The Episodic Memory Profile in Autism Spectrum Disorder: A Bayesian Meta-Analysis. *Neuropsychology Review*, *32*(2), 316–351. https://doi.org/10.1007/s11065-021-09493-5

Hauck, M., Fein, D., Maltby, N., Waterhouse, L., & Feinstein, C. (1998). Memory for Faces in Children with Autism. *Child Neuropsychology (Neuropsychology, Development and Cognition: Section C)*, *4*(3), 187–198. https://doi.org/10.1076/chin.4.3.187.3174

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Jacob, S., Wolff, J. J., Steinbach, M. S., Doyle, C. B., Kumar, V., & Elison, J. T. (2019). Neurodevelopmental heterogeneity and computational approaches for understanding autism. *Translational Psychiatry*, *9*(1), 63. https://doi.org/10.1038/s41398-019-0390-0

Jiang, C. C., Lin, L. S., Long, S., Ke, X. Y., Fukunaga, K., Lu, Y. M., & Han, F. (2022). Signalling pathways in autism spectrum disorder: mechanisms and therapeutic implications. *Signal Transduction and Targeted Therapy*, *7*(1), 229. https://doi.org/10.1038/s41392-022-01081-0

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

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Defining and Quantifying the Social Phenotype in Autism. *American Journal of Psychiatry*, *159*(6), 895–908. https://doi.org/10.1176/appi.ajp.159.6.895

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, *21*(9), 1148–1160. https://doi.org/10.1038/s41593-018-0210-5

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(NOV), 1–28. https://doi.org/10.3389/neuro.06.004.2008

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *ImageNet Classification with Deep Convolutional Neural Networks*, *1*, 1097–1105.

Krossa, L., Azimi, T., Lieberman, J., & Chen, L. (2023). Limited Neural Capacity and Hyper-Excitability Affect Quantity Processing: A Computational Account. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lind, S. E., & Bowler, D. M. (2010). Episodic memory and episodic future thinking in adults with autism. *Journal of Abnormal Psychology*, *119*(4), 896–905. https://doi.org/10.1037/a0020631

Liu, C., Jia, S., Liu, H., Zhao, X., Li, C. T., Xu, B., & Zhang, T. (2025). Recurrent neural networks with transient trajectory explain working-memory encoding mechanisms. Communications Biology, 8, Article 137.

Liu, J., Chang, H., Abrams, D. A., Kang, J. B., Chen, L., Rosenberg-Lee, M., & Menon, V. (2023). Atypical cognitive training-induced learning and brain plasticity and their relation to insistence on sameness in children with autism. *Elife*, *12*, e86035.

Liu, J., Chen, L., Chang, H., Rudoler, J., Al-Zughoul, A. B., Kang, J. B., Abrams, D. A., & Menon, V. (2023). Replicable Patterns of Memory Impairments in Children With Autism and Their Links to Hyperconnected Brain Circuits. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *8*(11), 1113–1123. https://doi.org/https://doi.org/10.1016/j.bpsc.2023.05.002

Lynch, C. J., Uddin, L. Q., Supekar, K., Khouzam, A., Phillips, J., & Menon, V. (2013). Default mode network in childhood autism: Posteromedial cortex heterogeneity and relationship with social deficits. *Biological Psychiatry*, *74*(3), 212–219. https://doi.org/10.1016/j.biopsych.2012.12.013

Lynn, A. C., Padmanabhan, A., Simmonds, D., Foran, W., Hallquist, M. N., Luna, B., & O'Hearn, K. (2018). Functional connectivity differences in autism during face and car recognition: underconnectivity and atypical age-related changes. *Developmental Science*, *21*(1), e12508. https://doi.org/10.1111/desc.12508

Maass, W. (1997). Networks of spiking neurons: The third generation of neural network models. Neural Networks, 10(9), 1659–1671.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nat Rev Neurosci*, *4*(4), 310–322. http://dx.doi.org/10.1038/nrn1076

Mehrer, J., Spoerer, C. J., Kriegeskorte, N., & Kietzmann, T. C. (2020). Individual differences among deep neural network models. *Nature Communications*, *11*(1), 2020.01.08.898288. https://doi.org/10.1038/s41467-020-19632-w

Minio-Paluello, I., Porciello, G., Pascual-Leone, A., & Baron-Cohen, S. (2020). Face individual identity recognition: a potential endophenotype in autism. *Molecular Autism 2020 11:1*, *11*(1), 1–16. https://doi.org/10.1186/S13229-020-00371-0

Mundy, P., & Stella, J. (2000). Joint attention, social orienting, and nonverbal communication in autism. In *Autism spectrum disorders: A transactional developmental perspective* (pp. 55–77). http://openurl.library.uiuc.edu/sfxlcl3?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:book&genre=bookitem&sid=ProQ:PsycINFO&atitle=Joint+attention%252C+social+orienting%252C+and+nonverbal+communication+in+autism.&title=Autism+spectrum+disorders%253

O'Hearn, K., Larsen, B., Fedor, J., Luna, B., & Lynn, A. (2020). Representational similarity analysis reveals atypical age-related changes in brain regions supporting face and car recognition in autism. *NeuroImage*, *209*, 116322. https://doi.org/10.1016/J.NEUROIMAGE.2019.116322

Pallett, P. M., Cohen, S. J., & Dobkins, K. R. (2014). Face and object discrimination in autism, and relationship to IQ and age. *Journal of Autism and Developmental Disorders*, *44*(5), 1039–1054. https://doi.org/10.1007/s10803-013-1955-z

Park, W. J., Schauder, K. B., Zhang, R., Bennetto, L., & Tadin, D. (2017). High internal noise and poor external noise filtering characterize perception in autism spectrum disorder. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-17676-5

Pierce, K. (2001). Face processing occurs outside the fusiform `face area' in autism: evidence from functional MRI. *Brain*, *124*(10), 2059–2073. https://doi.org/10.1093/brain/124.10.2059

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

Qian, N., & Lipkin, R. M. (2011). A learning-style theory for understanding autistic behaviors. *Frontiers in Human Neuroscience*, *5*, 77.

Ring, H., Woodbury-Smith, M., Watson, P., Wheelwright, S., & Baron-Cohen, S. (2008). Clinical heterogeneity among people with high functioning autism spectrum conditions: evidence favouring a continuous severity gradient. *Behavioral and Brain Functions*, *4*(1), 11.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, *111*(1), 205–235. https://doi.org/10.1037/0033-295x.111.1.205

Rubenstein, J. L. R., Merzenich, M. M., Rubenstein J L R, Merzenich, M. M., Rubenstein, J. L. R., & Merzenich, M. M. (2010). Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain & Behavior*, *2*(5), 255–267. https://doi.org/10.1046/j.1601-183X.2003.00037.x

Schultz, R. T. (2005). Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area. *Int. J. Devl Neuroscience*, *23*, 125–141. file:///Users/Megan/Documents/Papers2/Articles/2005/Schultz/Developmental deficits in social perception in autism \nthe role of the amygdala and fusiform face area 2005 Schultz.pdf\npapers2://publication/doi/10.1016/j.ijdevneu.2004.12.012

Seidenberg, M. S. (2005). Connectionist Models of Word Reading. *Current Directions in Psychological Science*, *14*(5), 238–243.

South, M., Stephenson, K. G., Nielson, C. A., Maisel, M., Top, D. N., & Kirwan, C. B. (2015). Overactive Pattern Separation Memory Associated with Negative Emotionality in Adults Diagnosed with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *45*(11), 3458–3467. https://doi.org/10.1007/s10803-015-2547-x

Stantić, M., Ichijo, E., Catmur, C., & Bird, G. (2022). Face memory and face perception in autism. *Autism*, *26*(1), 276–280. https://doi.org/10.1177/13623613211027685

Suri, K., Lewis, M., Minar, N., Willson, E., & Ace, J. (2021). Face Memory Deficits in Children and Adolescents with Autism Spectrum Disorder. *Journal of Psychopathology and Behavioral Assessment*, *43*(1), 108–118. https://doi.org/10.1007/S10862-020-09840-5/TABLES/3

Vilidaite, G., Yu, M., & Baker, D. H. (2017). Internal noise estimates correlate with autistic traits. *Autism Research*, *10*(8), 1384–1391. https://doi.org/10.1002/aur.1781

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neuroscience and Biobehavioral Reviews*, *36*(3), 1060–1084. https://doi.org/10.1016/j.neubiorev.2011.12.008

Williams, D. L., Goldstein, G., & Minshew, N. J. (2006). The profile of memory function in children with autism. *Neuropsychology*, *20*(1), 21–29. https://doi.org/2006-01440-003 [pii]\r10.1037/0894-4105.20.1.21 [doi]

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.

Xu, D., et al.(2024). BrainCog: A spiking neural network based, brain-inspired cognitive intelligence engine for brain-like AI. Engineering, 21, 109–124.

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

Yizhar, O., Fenno, L. E., Prigge, M., Schneider, F., Davidson, T. J., Ogshea, D. J., Sohal, V. S., Goshen, I., Finkelstein, J., Paz, J. T., Stehfest, K., Fudim, R., Ramakrishnan, C., Huguenard, J. R., Hegemann, P., & Deisseroth, K. (2011). Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature*, *477*(7363), 171–178. https://doi.org/10.1038/nature10360