2018 Joint IEEE 8th International Conference on Development and Learning
and Epigenetic Robotics (ICDL-EpiRob)
Waseda University, Tokyo, Japan, Sept. 16-20, 2018

Understanding the cognitive mechanisms underlying autistic behavior: a recurrent neural network study

Anja Philippsen, Yukie Nagai
Center for Information and Neural Networks (CiNet),
National Institute of Information and Communications Technology (NICT),
Osaka, Japan
{anja, yukie}@nict.go.jp

***Abstract*—People with autism spectrum disorder are suggested to exhibit atypical perception and differences in cognitive processing. In behavioral studies, however, such differences are often difficult to verify. Apparently, differences in cognitive processing do not always cause an impairment of behavior. To investigate how such a mismatch between cognitive and behavioral level could be explained, we model and evaluate the process of learning to imitate using recurrent neural networks. We systematically adjust learning parameters of the network which are linked to the precision of learning, a factor that might differ between individuals with autism and typically developed individuals. We evaluate the trained networks in terms of task performance (behavioral level) as well as in terms of the structure of the internal representation that emerges during learning (cognitive level). Our findings demonstrate that comparable behavioral network output can be caused by different internal network representations. A less well structured internal representation does not necessarily result in a decline in performance, but can also be associated with good imitation performance. Additionally, we find evidence that well structured internal representations in our setting emerge with an appropriate integration of top-down predictions and bottom-up information processing, a finding which integrates well with theories from developmental psychology.**

I. Introduction

Autism spectrum disorder (ASD) is a prevalent developmental disorder that is characterized by difficulties in social interaction [1] and atypical perception [2], [3]. Although autism is considered as a neurological disorder [4], little is known about the underlying neurological causes [5]. Diagnosis, therefore, occurs based on behavioral observations [1] despite the large variety of symptoms in people with autism.

In search of cognitive mechanisms that could explain the symptoms of autism in a unified way, recently many studies focus on the predictive coding hypothesis [6]–[8] which interprets our perceptual understanding and our interaction with the world as the result of prediction error minimization. Following the Bayesian formulation of predictive coding, our brain integrates top-down knowledge (such as expectations, contextual information etc.) with bottom-up sensory perceptions. ASD might be caused by an imbalance between these two processing pathways [9], [10]. In these terms, symptoms of autism could be explained by a reduced usage of top-down knowledge [9] or by an increased precision of the bottom-up sensory input [11].

To evaluate the plausibility of such theories and to propose concrete cognitive mechanisms that can cause typical symptoms of autism, computational models can be applied. In particular, recurrent neural networks have proved useful to demonstrate how certain network parameters can be altered to replicate typical behavioral patterns observed in subjects with disorders such as autism or schizophrenia [12], [13]. In [12], Yamashita and Tani employ a recurrent neural network with two layers working on different timescales. They demonstrate that disconnectivity between these layers causes symptoms in the model which can be related to symptoms of schizophrenia. Similarly, Idei et al. [13] could replicate repetitive behavior patterns of ASD subjects in a recurrent neural network model by adjusting the sensitivity of the network to the estimated variance of the external signal. As suggested in [10], such an altered confidence could increase or decrease the prediction error and impair predictive learning.

Whereas these studies successfully identify parameters which can alter the predictions of the recurrent neural network, they take only the behavioral level into account which captures the true nature of the disorder only partially. A comparison of behavioral and neuroimaging studies of ASD indicates that atypical cognitive processing does not necessarily cause behavioral differences [14], in particular, because many people with autism implement strategies to cope with difficulties in their everyday life by adopting behavioral patterns which they discovered to be appropriate [14], [15]. Therefore, despite differences in cognitive processing, the observed behavior, especially in adult subjects, can be very similar to that of typically developed individuals.

Another shortcoming of studies such as [12], [13] is that they make use of pretrained network models and change the parameters only during execution time. The characterization of ASD as a *developmental* disorder, however, suggests that cognitive differences are already present during development and shape the development of representation structure in the brain.

In this study we address these shortcomings. Specifically, we demonstrate in a recurrent neural network model that differences in cognitive development can cause differences in internal representation structure, whereas the network's

978-1-5386-6110-9/18/$31.00 ©2018 IEEE

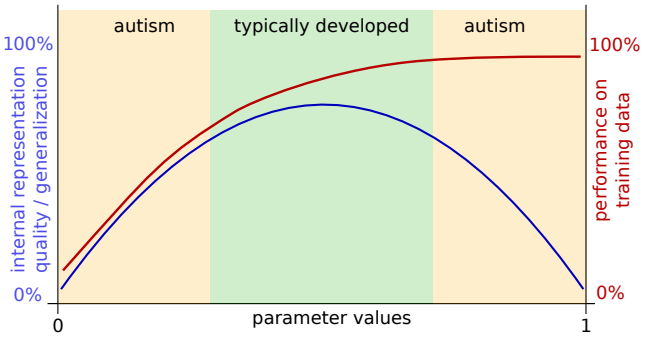

Fig. 1. Schematic view on how network performance and the quality of the internal network representation structure may have different optimal parameter values.

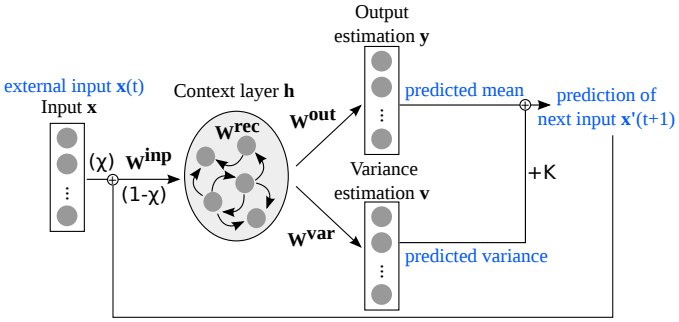

Fig. 2. Stochastic Continuous-Time Recurrent Neural Network: A recurrent layer of context neurons processes the external input and predicts the next time step by estimating mean and variance of the external signal.

performance on the training data remains intact.

We evaluate independently the effect of two parameters on the learning process. The first parameter, *external contribution*, expresses the network's sensitivity to external input. People with autism are often considered to exhibit very precise perception which might coincide with a relatively smaller usage of their own top-down predictions [9], [16]. The second parameter is the *aberrant precision* parameter introduced in [13] which causes the network to over- or underestimate the noise in the environment, similarly to how people with autism are hypothesized to overestimate the volatility of the environment [10], [17].

We adjust these two parameters during the learning process and evaluate the effect that the parameters have on the achieved performance (behavioral level) and on the quality of the emerged internal network representations (cognitive level). A mismatch between these two levels could look like it is depicted in Fig. 1. In this example, optimizing the parameter for performance on the training data leads to overtraining which might be reflected in a less well structured internal representation. Poor internal representation structure, therefore, can be associated with good task performance (right side of Fig. 1) as well as with poor task performance (left side of Fig. 1). Such a finding could make it plausible that the performance for a specific task can be intact in subjects with ASD despite differences in cognitive processing.

II. A Model of Prediction: Stochastic Continuous-Time Recurrent Neural Network

Stochastic Continuous-Time Recurrent Neural Networks (S-CTRNNs) are recurrent neural network models that can learn time series by mapping an input to an output signal [18], [19]. In contrast to standard recurrent neural networks, S-CTRNNs do not only estimate the average value of the target signal, but also learn to estimate the signal's time-varying variance. The weight update via backpropagation is not only based on the distance of the predicted to the actual signal (prediction error), but makes use of the *scaled* prediction error: by inversely weighting the error of the prediction with the estimated variance, the influence of the prediction error is reduced in case of high variance estimation. As a result, a more stable learning of noisy training signals becomes possible [19].

A. Network overview and training procedure

A schematic view of a S-CTRNN is presented in Fig. 2. The network is trained to predict time series $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_t, \ldots, \boldsymbol{x}_T$ by estimating the next time step $\boldsymbol{x}_{t+1}$ from the current step $\boldsymbol{x}_t$ and from the history of computations that is captured in the recurrent context layer of the network.

Analogously to the definitions in [19], the internal states of the neurons in the recurrent network layer $u^{\text{rec}}_{t,i}$ are updated by integrating the internal states from the previous time step with the current input according to a time scale parameter τ:

$$u^{\text{rec}}_{t,i} = (1-\frac{1}{\tau})u^{\text{rec}}_{t-1,i} + \frac{1}{\tau}(\sum_{j=1}^{I} w^{\text{inp}}_{ij} x_{t,j} + \sum_{j=1}^{C} w^{\text{rec}}_{ij} h_{t-1,j}), \quad (1)$$

where I and C are the dimensions of input and context layer, respectively, and w_{ij} denotes the incoming synaptic weight of neuron i from neuron j. The variable $h_{t,i} = \tanh(u^{\text{rec}}_{t,i})$ denotes activation values of the hidden context neurons.

The internal states of the neurons in the output layers are computed directly from the network state and the corresponding output weights:

$$u^{\text{out}}_{t,i} = \sum_{j=1}^{C} w^{\text{out}}_{ij} h_{t,j}, \quad u^{\text{var}}_{t,i} = \sum_{j=1}^{C} w^{\text{var}}_{ij} h_{t,j}. \quad (2)$$

The formulas for computing the output estimation $\boldsymbol{y}$ and variance estimation $\boldsymbol{v}$ are:

$$y_{t,i} = \tanh(u^{\text{out}}_{t,i}), \quad (3)$$

$$v_{t,i} = \exp(u^{\text{var}}_{t,i}). \quad (4)$$

The weights w_{ij} of the network are adjusted in order to maximize the likelihood that the estimated mean and variance values describe the true time series $\boldsymbol{x}$. Concretely, in each training epoch, we minimize the negative log likelihood of the predicted training signal, defined as:

$$-lnL = \sum_{t=1}^{T}\sum_{i=1}^{O}(ln(2\pi v_{t,i}) + \frac{(x_{t+1,i} - y_{t,i})^2}{2v_{t,i}}), \quad (5)$$

where O denotes the output dimensionality.

Training proceeds by presenting training trajectories to the network, computing the likelihood in (Eq. 5) and optimizing all network weights accordingly [20]. During training, for different trajectories, different initial states of the context layer neurons are set, such that the network can differentiate different trajectories. The initial state for each neuron i is initialized with $u^{\text{rec}}_{0,i} = 0$, and optimized during training according to the likelihood function (Eq. 5). As with this function alone, the initial states would constantly diverge (to achieve better separation), a second optimization function is employed additionally that keeps the variance of the initial states close to some predefined variance value (cf. [19]).

We implemented the S-CTRNN via the CHAINER framework [21]. The network's context layer consists of 70 neurons and as time scale parameter we chose $\tau = 2$.

B. Learning parameters

1) External contribution: The external contribution parameter χ expresses the relative amount of information from the external signal that the network uses for updating its internal representation (cf. Fig. 2). If this parameter is varied during execution time, two different modes of execution can be realized. If $\chi_{test} = 1$, the network relies solely on external input, ignoring its own output estimation. The network, thus, *reactively* follows the external signal without regard to its own estimation. If smaller values of χ_{test} are employed, the network combines the external input with its own estimation. Relying completely on its own estimation ($\chi_{test} = 0$), the network employs *proactive* behavior. Sufficiently trained networks can proactively generate a learned trajectory if the context layer activations are initialized according to the desired trajectory.

Adjusting χ during the learning process, we can control how much the system focuses on the external signal during training. As with $\chi_{train} = 0$ the network ignores external input, and nothing new can be learned, a minimum value of $\chi_{train} = 0.1$ is assumed in this study. Specifically, we evaluate values of $\chi_{train} \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$.

The integrated target signal $\boldsymbol{x}'$ that forms the input to the context layer (cf. Fig. 2) is generated by combining the external target signal $\boldsymbol{x}$ with the estimated signal which is determined by adding Gaussian noise with variance $\boldsymbol{v}$ to the averaged output $\boldsymbol{y}$ (the estimated signal, thus, mimics the noise property of the original target trajectory):

$$\boldsymbol{x}'_{t+1} = \chi \boldsymbol{x}_{t+1} + (1 - \chi)\mathcal{N}(\boldsymbol{y}_t, \boldsymbol{v}_t) \quad (6)$$

When applying a value of $\chi < 1$ during training, not only the input that is provided to the context layer changes, but also the computation of the prediction error. Instead of comparing the predicted signal to the external signal $\boldsymbol{x}_{t+1}$ in (Eq. 5), the estimated trajectory is compared to the integrated trajectory $\boldsymbol{x}'_{t+1}$.

With $\chi = 1$, the learner very precisely tries to reproduce the external signal. A reduced χ reflects a trade-off between the raw external (bottom-up) information and the system's own (top-down) predictions. Similarly to how an integration of these two processing pathways seems to contribute to typical development [9], a medium level of χ might be beneficial for balancing performance and a well structured internal representation.

2) Aberrant precision: The aberrant precision parameter K [13] alters the estimation of variance output in (Eq. 4) to the following formula:

$$v_{t,i} = \exp(u^{\text{var}}_{t,i} + K). \quad (7)$$

With $K = 0$, the network is trained as usual. Negative values of K reduce the estimated variance and, thus, lead to a general underestimation of the signal variance. Positive values of K lead to an overestimation of variance.

In [13], both, an over- and an underestimation of the prediction error reduced the network's ability to flexibly switch between different behaviors according to the situation. In contrast to [13], we alter the parameter already during the learning process. Negative values of K increase the influence of the prediction error which can be assumed to accelerate weight updates in the beginning of learning, whereas positive values reduce the prediction error, which could result in slower learning progress. We evaluate values of $K \in \{-8, -4, -2, 0, 2, 4, 8\}$.

A reduced or increased confidence in their own predictions might be a cause for failure in prediction in ASD subjects [10]. The parameter K reflects how strong the network weights its estimated variance and, therefore, implements aberrant precision during the learning process.

C. Imitation task

The learning task we consider is the imitation of two-dimensional trajectories generated from Lissajous curves with additive Gaussian noise (also used in [19]). Eight different trajectories are used: four ellipses and four "eight" shapes located at four different positions of the two-dimensional space as depicted in Fig. 3. The sampling rate is chosen such that a single execution of the trajectory consists of 25 time steps.

The variance of Gaussian noise differs for different trajectories such that $\sigma_{\text{ellipse1}} = \sigma_{\text{ellipse2}} = 0.001$, $\sigma_{\text{ellipse3}} = \sigma_{\text{ellipse4}} = 0.003$, $\sigma_{\text{eight1}} = \sigma_{\text{eight2}} = 0.005$ and $\sigma_{\text{eight3}} = \sigma_{\text{eight4}} = 0.007$. With this design, the network may internally categorize the different trajectories by different criteria: the shape of the trajectory, the position in space or the level of signal noise.

III. Evaluation of cognitive processing

Network training is not only affected by the learning parameters, but also by the initial network connection weights. To reduce initial weight influence on the evaluation, the same initial weights are used for all parameter conditions within one trial. We perform 10 independent trials using different

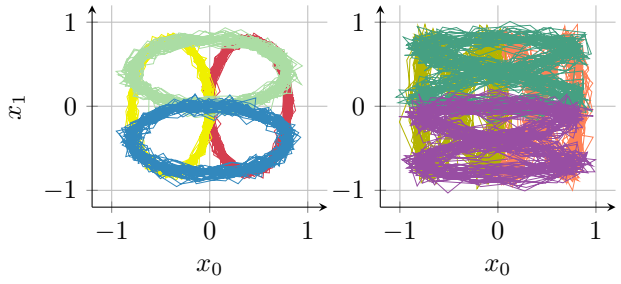

Fig. 3. The 8 two-dimensional training trajectories generated from Lissajous curves with increasing levels of Gaussian-distributed noise, plotted in two separate plots for the sake of clarity.

initial weights. In each trial, for each parameter value in the ranges defined in Sec. II-B, the network is trained for 2000 epochs. In each epoch, the network is presented with one training sequence (containing 40 repetitions of the trajectory which results in 1000 time steps) of each of the eight training trajectories. If χ_{train} is varied, $K = 0$ is used, whereas in the experiments with different values of K, χ_{train} is set to 1.

A. Task-specific performance

The task-specific performance of a network is evaluated in terms of its imitation capability. The network imitates a trajectory by setting the context activations to the initial state which best predicts the desired trajectory. Then, the network performs proactive ($\chi_{test} = 0$) generation, using its own output as new input. For the prediction of the target trajectory during training, the estimated variance of the trajectory was included for signal integration (Eq. 6). For imitation, we assume proactive generation of the target trajectory without replication of the noise properties. This equation, thus, simplifies to $\boldsymbol{x}'_{t+1} = \boldsymbol{y}_t$.

The network's performance is evaluated by letting it reproduce freshly generated target trajectories of 75 time steps (i.e. 3 repetitions) for each target shape. Fig. 4 shows the network's prediction errors (averaged over the eight target trajectories) over the course of training. The left column corresponds to the mean square error, computed directly between the estimated output and the target trajectory.

For most parameter values, the trained networks are able to well imitate the target trajectories after around 1000 epochs. For the χ_{train} parameter (Fig. 4, top), good performance is achieved with $\chi_{train} \geq 0.5$. As expected, with low values of χ_{train}, learning is impaired because the network strongly concentrates on its own predictions, ignoring the external signal.

For the K parameter (Fig. 4, bottom), a U-shaped performance can be observed: extreme deviations of estimated variance of $K = -8$ or $K = 8$ lead to a lower performance, but medium values perform well. Slightly negative values seem to even increase learning speed: The underestimation of variance increases the influence of the prediction error and causes higher weight updates.

The right column depicts the prediction errors scaled by the network's estimated variance which provides comparable

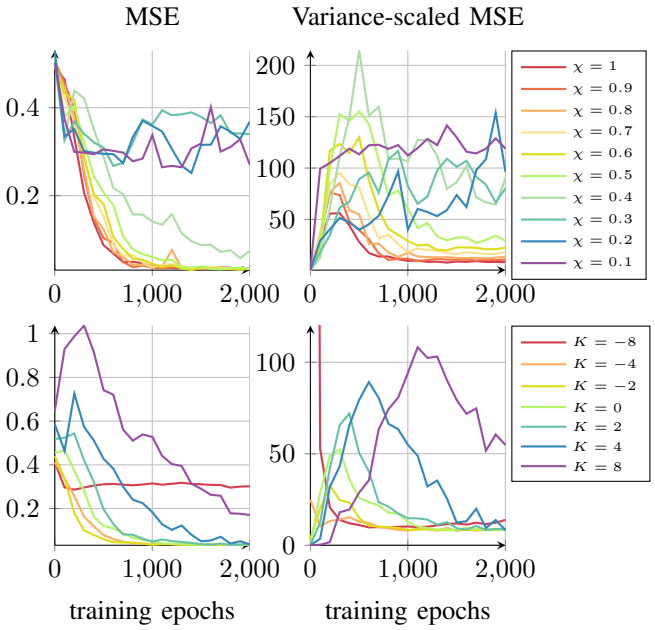

Fig. 4. Average prediction error (left: mean square error, right: variance-scaled mean square error) over all trials of predicted trajectories during proactive generation ($\chi_{test} = 0$) over the course of training, for different values of χ_{train} (top) and K (bottom), evaluated every 100 training epochs.

results. The lower error values in the first learning epochs can be attributed to a high variance estimation in the beginning of learning.

B. Qualitative evaluation of internal representation

Which internal representations do the networks acquire by learning to achieve this task? The internal representation of a recurrent neural network is coded in the dynamics of its context layer which can be assessed by observing the time course of context neuron activations while performing a task. We evaluate the time course of activations in the trained networks (after 2000 epochs) during the above described proactive imitation. By applying principal component analysis (PCA) on the network activations during trajectory generation, the original 70-dimensional context activation space can be projected to a time course in a lower dimensional representation [19]. Fig. 5 depicts the first two principal components of the activation patterns emerging with different parameter values for example trials.

For both parameters, small parameter values lead to unstructured dynamics in the network's context layer. These networks are not able to properly achieve the task and also could not build up an appropriate internal representation. The most structured patterns occur for medium parameter values, whereas high parameter values rather cause a stronger overlap in the context activations.

In general, it can be observed that networks seem to preferably represent patterns according to their position in space. Circles and eight shapes which are located at the same positions in task space (coded with similar colors)

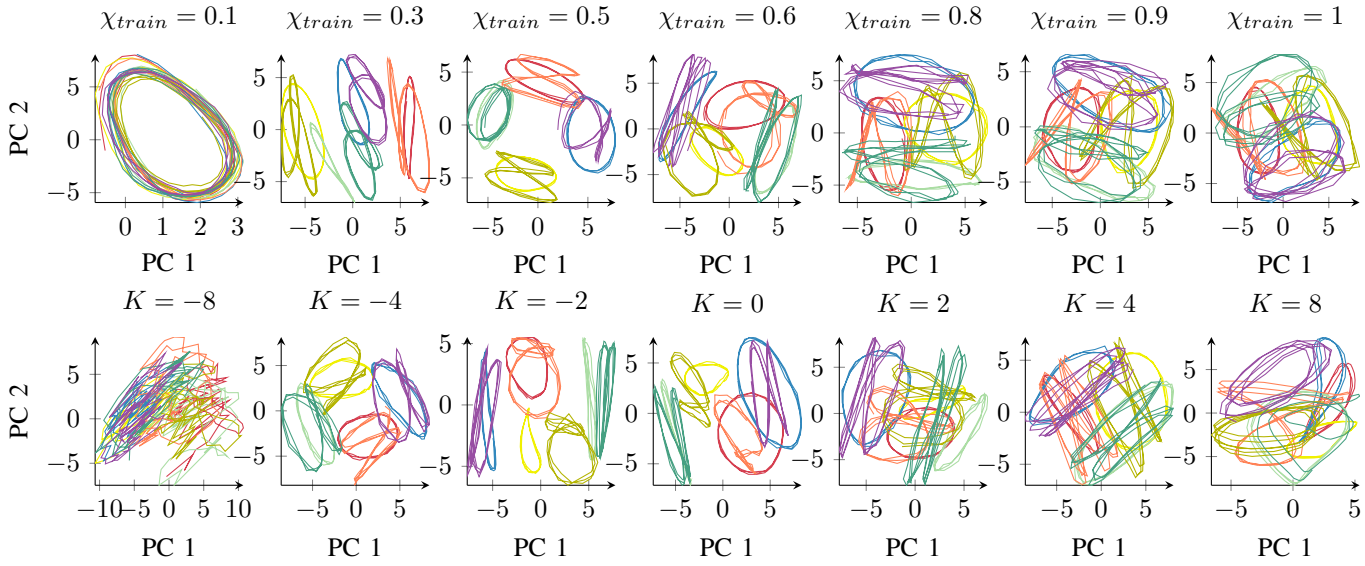

Fig. 5. Two-dimensional PCA of network activations of one trial (using the same initial network weights) for proactively following the target trajectories for different values of χ_{train} (top) or K (bottom) after training for 2000 epochs. Legend: see Fig. 3.

are represented by spatially similar activation patterns in the context neurons. Positions, thus, seem to be coded as a kind of bias in the context activations. Similar trajectory shapes seem to coincide with similar time courses of activation patterns. In contrast, the amount of noise present in the external signal has no apparent influence on the context activation structure.

C. Quantitative evaluation of internal representation

In the two-dimensional PC space, we found a preference for medium values of the parameters K and χ_{train}. But does this also hold for the full 70-dimensional activation space?

The qualitative evaluation suggests that context activation patterns primarily represent the position of the target trajectory in task space. In these terms, a well structured internal network representation can be quantitatively determined by comparing two different distance measures. We define *inner distances* as the distances between activation patterns of trajectories located at the same task space position. *Outer distances* express the distances of activation patterns of different-position trajectories. A good structure is reflected by low inner distances and high outer distances, thus, by a small quotient of inner to outer distances. Distances between two context neuron activation patterns are assessed via dynamic time warping [22] over the activation courses of the individual neurons.

Fig. 6 shows the mean inner and outer distances for different parameter values averaged over all trials (bold lines) and for all individual trials (thin lines). The inner-to-outer quotient is smaller in the case of better separation.

For the parameter χ_{train}, the best separation between same-position patterns is achieved with $\chi_{train} = 0.5$. Small values of χ_{train} clearly lead to a worse separation. Larger values of $\chi_{train} > 0.5$ also seem to coincide with weaker separation. Although the internal representations do not significantly differ

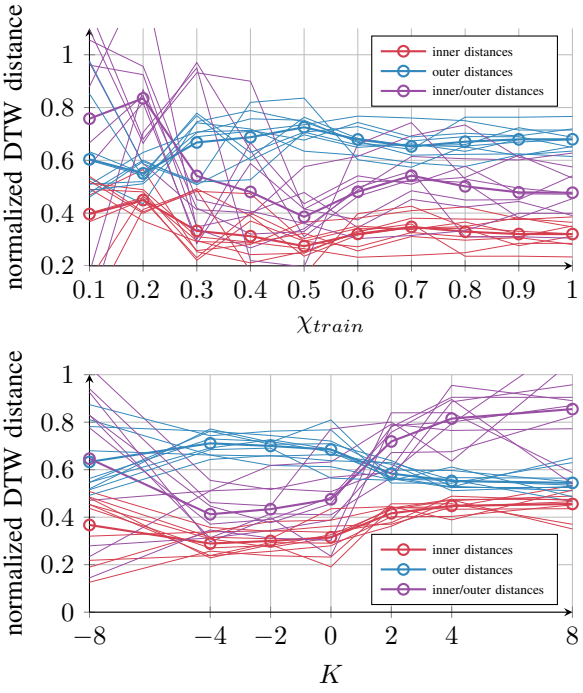

Fig. 6. Inner and outer distances for different parameter values of χ_{train} (top) and K (bottom). Bold lines mark the average over all 10 trials, thin lines represent the individual trials.

within this parameter range, we can observe a trend that suggests better separation for medium values of χ_{train} which matches the qualitative observation (Fig. 5).

For the parameter K, a more pronounced effect can be observed. In particular, although parameter values between $K = -4$ and $K = 4$ lead to equally good performance on

the behavioral level (cf. Fig. 4), we can observe a drastic difference in the internal representation structures. Similar to our qualitative observation (Fig. 5), the best separation is achieved with $K = -4$, -2 or 0.

IV. Discussion

Fig. 7 summarizes the findings from this study, depicting task performance (without and with integration of estimated variance) and representation quality after 2000 epochs. To make the results comparable to the schematic view in Fig. 1, the results are normalized to range $[0, 1]$ across parameter conditions and inverted.

Our findings demonstrate that good performance on the training data can be achieved even with suboptimal internal representation structure. Good performance is achieved for all values between $K = -4$ and $K = 4$, although values $K > 0$ obtain a significantly poorer representation structure. The gap between representation and performance level can be observed especially well for parameter values $K = 2$ and $K = 4$. Also for χ_{train}, a better task-specific separation in the context activations appears to be acquired for parameter values which achieve sub-optimal task performance, although the effect for this parameter is weaker than for K.

Apparently, no one-to-one correspondence between task-specific performance and internal representation structure exists. This finding supports the hypothesis that ASD might be characterized by a suboptimal internal representation, but does not necessarily cause a performance decrease [14].

Especially the results for χ_{train} are interesting as this parameter reflects how precisely the network tries to reproduce the external signal. Medium parameter values around $\chi_{train} = 0.5$ seem to better promote the emergence of a well structured internal representation compared to extreme parameter values. Suboptimal internal representation structure, thus, might emerge if someone very precisely tries to reproduce the bottom-up information (right side of Fig. 1) or too strongly relies on its own predictions (left side of Fig. 1). This supports the theory suggested by Pellicano and Burr [9] who attribute autism to a failure of top-down and bottom-up integration. However, as the effect is not very pronounced here, future research is required to validate this trend.

Studies with ASD subjects often point out problems in generalization and category formation [23]–[25]. In this study, we evaluate the network's performance by testing whether it is able to reproduce the training data (a common way to evaluate recurrent neural networks [12]). The network's *generalization* performance, in contrast, might be related to the quality of the network's internal representation. Poorly structured internal representations in our parameter settings are achieved with extreme values of χ_{train} or K, which are parameter conditions which express strong or insufficient precision during learning. Aberrant precision is often associated with autism [10], therefore, our results can be considered to be in line with these studies.

One could argue that the effects we found are only an effect of training duration as both parameters in a way affect the learning speed of the network. A larger value of χ_{train} calls for more precision and, thus, increases the prediction error which results in stronger weight updates. The weaker separation of same-place patterns in the context activations, thus, might be a result of faster learning. The parameter K affects the learning rate, as well: a high value of K leads to an overestimation of variance and a smaller prediction error. Learning, thus, is slowed down with high values of K. Still, we observe that a higher value of K causes a weaker separation in the internal representation. Due to this mismatch, learning rate alone does not provide a coherent explanation of how these parameters affect the structuring of the internal representation. Apart from their influence on the learning rate, the parameters appear to affect the learning process in a more complex way which requires further investigation.

All in all, the findings in this study support our endeavour to evaluate behavior in people with autism not only at a behavioral level, but also at the level of cognitive processing. In order to assess human cognitive processes and for making them quantitatively assessable, we plan to develop a *cognitive mirroring* system [26]: A robot should learn an internal model of the human's cognitive processing by interacting with a human and imitating the human's behavior patterns. Whereas not all aspects of cognitive processing might be measurable in terms of behavior, by trying to match the human's behavior the network can explore possible parameter configurations which might describe the underlying cognitive processes which are otherwise difficult to estimate. Unlike the human's cognitive state, the robot's acquired model is fully observable and can be quantitatively evaluated. Such a cognitive mirroring system could provide valuable hints for a therapist, indicating which internal mechanisms might cause an observed behavior, suggesting pathways on how to support a person with autism, for example, by implementing changes in the environment.

The results presented here constitute a first step toward this aim by suggesting how different learning mechanisms affect the network's internal representation. The observation that similar behavior can be caused by different internal representations suggests that in addition to performance monitoring, close attention should be paid to the emerging internal network representations.

V. Conclusion

People with ASD exhibit differences in cognitive processing although their behavior often seems to be similar to that of typically developed subjects. We demonstrate a similar effect in a recurrent neural network model by adjusting different learning parameters and evaluating the trained networks not only in terms of performance but also in terms of the acquired internal representation structure. We found a discrepancy between performance and representation quality, indicating that an evaluation of the behavioral level alone is insufficient for investigating the causes of autism. Rather, it is crucial to gain a better understanding of the underlying cognitive processes.

In this study we use only a very simple, artificial task, therefore, we want to extend the evaluation to more realistic

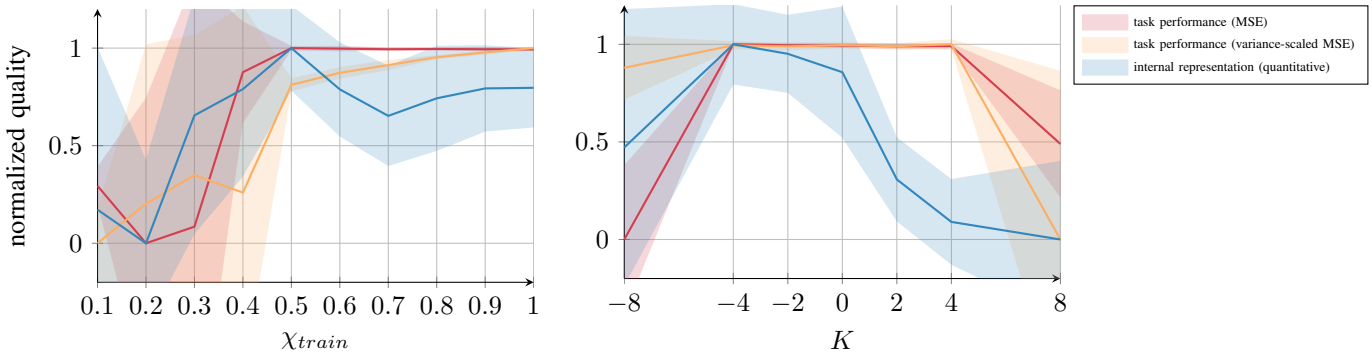

Fig. 7. Summary of results from this study for parameters χ_{train} (left) and K (right). Mean and standard deviation of the results are normalized and inverted to make them comparable to Fig. 1.

drawing tasks in the future. Also, considering the large differences between the human perceptual and cognitive system and the proposed model, cognitive mechanisms of this model are of course not directly transferable to neurocognitive theories. However, they can still provide a measure of cognitive processing that might have analogies to similar processes in humans.

ACKNOWLEDGEMENT

This work was supported by JST CREST "Cognitive Mirroring: Assisting people with developmental disorders by means of self-understanding and social sharing of cognitive processes" (Grant Number: JPMJCR16E2), Japan.

REFERENCES

[1] C. Lord, S. Risi, L. Lambrecht, E. H. Cook, B. L. Leventhal, P. C. DiLavore, A. Pickles, and M. Rutter, "The autism diagnostic observation schedule–generic: A standard measure of social and communication deficits associated with the spectrum of autism," *Journal of autism and developmental disorders*, vol. 30, no. 3, pp. 205–223, 2000.

[2] O. Bogdashina, *Sensory perceptual issues in autism and asperger syndrome: different sensory experiences-different perceptual worlds*. Jessica Kingsley Publishers, 2016.

[3] R. P. Lawson, J. Aylward, S. White, and G. Rees, "A striking reduction of simple loudness adaptation in autism," *Scientific reports*, vol. 5, p. 16157, 2015.

[4] N. J. Minshew and D. L. Williams, "The new neurobiology of autism: cortex, connectivity, and neuronal organization," *Archives of neurology*, vol. 64, no. 7, pp. 945–950, 2007.

[5] E. Anagnostou and M. J. Taylor, "Review of neuroimaging in autism spectrum disorders: what have we learned and where we go from here," *Molecular autism*, vol. 2, no. 1, p. 4, 2011.

[6] R. P. Rao and D. H. Ballard, "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects," *Nature neuroscience*, vol. 2, no. 1, p. 79, 1999.

[7] J. Tani, "Learning to generate articulated behavior through the bottom-up and the top-down interaction processes," *Neural Networks*, vol. 16, no. 1, pp. 11–23, 2003.

[8] K. Friston, "The free-energy principle: a rough guide to the brain?" *Trends in cognitive sciences*, vol. 13, no. 7, pp. 293–301, 2009.

[9] E. Pellicano and D. Burr, "When the world becomes too real: a bayesian explanation of autistic perception," *Trends in cognitive sciences*, vol. 16, no. 10, pp. 504–510, 2012.

[10] R. P. Lawson, G. Rees, and K. J. Friston, "An aberrant precision account of autism," *Frontiers in human neuroscience*, vol. 8, p. 302, 2014.

[11] H. Haker, M. Schneebeli, and K. E. Stephan, "Can bayesian theories of autism spectrum disorder help improve clinical practice?" *Frontiers in psychiatry*, vol. 7, p. 107, 2016.

[12] Y. Yamashita and J. Tani, "Spontaneous prediction error generation in schizophrenia," *PLoS One*, vol. 7, no. 5, p. e37843, 2012.

[13] H. Idei, S. Murata, Y. Chen, Y. Yamashita, J. Tani, and T. Ogata, "Reduced behavioral flexibility by aberrant sensory precision in autism spectrum disorder: A neurorobotics experiment," in *2017 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, Sept 2017, pp. 271–276.

[14] M. B. Harms, A. Martin, and G. L. Wallace, "Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies," *Neuropsychology review*, vol. 20, no. 3, pp. 290–322, 2010.

[15] S. Kumagaya, "Tojisha-kenkyu of autism spectrum disorders," *Advanced Robotics*, vol. 29, no. 1, pp. 25–34, 2015.

[16] S. Van de Cruys, K. Evers, R. Van der Hallen, L. Van Eylen, B. Boets, L. de Wit, and J. Wagemans, "Precise minds in uncertain worlds: Predictive coding in autism." *Psychological review*, vol. 121, no. 4, p. 649, 2014.

[17] R. P. Lawson, C. Mathys, and G. Rees, "Adults with autism overestimate the volatility of the sensory environment," *Nature neuroscience*, vol. 20, no. 9, p. 1293, 2017.

[18] J. Namikawa, R. Nishimoto, H. Arie, and J. Tani, "Synthetic approach to understanding meta-level cognition of predictability in generating cooperative behavior," in *Advances in Cognitive Neurodynamics (III)*. Springer, 2013, pp. 615–621.

[19] S. Murata, J. Namikawa, H. Arie, S. Sugano, and J. Tani, "Learning to reproduce fluctuating time series by inferring their time-dependent stochastic properties: Application in robot learning via tutoring," *IEEE Transactions on Autonomous Mental Development*, vol. 5, no. 4, pp. 298–310, 2013.

[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[21] S. Tokui, K. Oono, S. Hido, and J. Clayton, "Chainer: a next-generation open source framework for deep learning," in *Proceedings of Workshop on Machine Learning Systems (LearningSys) in the 29th Annual Conference on Neural Information Processing Systems (NIPS)*, vol. 5, 2015.

[22] M. Müller, "Dynamic time warping," *Information retrieval for music and motion*, pp. 69–84, 2007.

[23] N. J. Minshew, J. Meyer, and G. Goldstein, "Abstract reasoning in autism: A disassociation between concept formation and concept identification." *Neuropsychology*, vol. 16, no. 3, p. 327, 2002.

[24] H. Z. Gastgeb, E. M. Dundas, N. J. Minshew, and M. S. Strauss, "Category formation in autism: can individuals with autism form categories and prototypes of dot patterns?" *Journal of autism and developmental disorders*, vol. 42, no. 8, pp. 1694–1704, 2012.

[25] A. Froehlich, J. Anderson, E. Bigler, J. Miller, N. Lange, M. DuBray, J. Cooperrider, A. Cariello, J. Nielsen, and J. Lainhart, "Intact prototype formation but impaired generalization in autism," *Research in autism spectrum disorders*, vol. 6, no. 2, pp. 921–930, 2012.

[26] Y. Nagai, "Cognitive mirroring: Assisting people with developmental disorders by means of self-understanding and social sharing of cognitive processes," *Seitai No Kagaku (in Japanese)*, vol. 69, no. 1, pp. 63–67, 2018.