Artificial Neural Networks trained on human cognitive data as network models of cognition in health and mental disorders

Oliver Frank, Daniel Durstewitz, Emanuel Schwarz, Georgia Koppe, Urs Braun
Hector Institute for Artificial Intelligence in Psychiatry - Central Institute of Mental Health Mannheim, Germany

Abstract

Cognitive functions are mental processes essential for goal-directed behavior. Impairments in these functions are common in psychiatric disorders and significantly impact quality of life. Artificial Neural Networks (ANNs), trained on cognitive test data from human individuals, offer a new model-based approach to study potential causal links between brain network structure, cognitive function and brain architecture.

In this study, we collected longitudinal cognitive data from healthy individuals and patients (schizophrenia, depression, autism spectrum disorder) to train individualized ANNs and analyse their emerging network properties. Our results show that ANNs can learn participants' behavior and, when initialized with suitable architectures, exhibit a balance of integration and segregation in their hidden layers, mirroring the brain's topological organization. Network topologies remain mostly robust across randomized training iterations, and topological marker distributions differ significantly (5 out of 6 comparisons (t-test), $p < .05$). Our findings suggest that ANNs trained on cognitive-behavioral data may serve as tools to understand (brain) network properties underlying human cognitive function in health and mental disorder.

Keywords: Machine Learning; Artificial Neural Network; Graph-theoretical Marker; Ecological Momentary Assessment; Mental Disorder

Introduction

Cognitive functions are mental processes crucial for goal-directed behavior. Impairments often occur in psychiatric disorders like schizophrenia, depression, and autism spectrum disorders, significantly affecting quality of life. Structurally, these functions rely on the coordinated activity of millions of neurons, which can be studied via methods like functional MRI (fMRI). These approaches have revealed that the brain exhibits both functional segregation — specialized clusters for task-specific processing — and global integration through central hubs enabling flexible cognition. A balance between these modes is believed to underpin cognitive capacity, while imbalances are linked to psychiatric conditions (Bassett & Sporns, 2017). However, due to limitations in neuroimaging, particularly in capturing dynamic and fine-grained network changes over time, such insights remain largely descriptive and lack causal precision. In parallel, ANNs have advanced considerably. Recent work suggests that ANNs trained on behavioral data can reflect organizational principles similar to those found in biological brains (Barak, 2017; Yang, Joglekar, Song, Newsome, & Wang, 2019).

In this study, we train ANNs on cognitive-behavioral data from healthy individuals and patients to analyze resulting network topologies using graph-theoretical methods. Our goals are twofold: to determine whether topological features of ANNs trained on individual datasets (1) can model fundamental organizational principles of human brain function, and (2) differ reliably between patients and controls — potentially enabling personalized diagnosis and tracking of cognitive change.

Experiment

This study employed a longitudinal design to collect the large dataset required for training the described ANNs. Four participants repeatedly completed a cognitive test battery consisting of twelve tasks on their smartphones over several months. The tasks covered cognitive domains such as decision-making, executive functioning, relational processing, and working memory (Figure 1), and were completed five times per week for 20 minutes each session. The task difficulty levels were individually adjusted based on the performance of each participant.

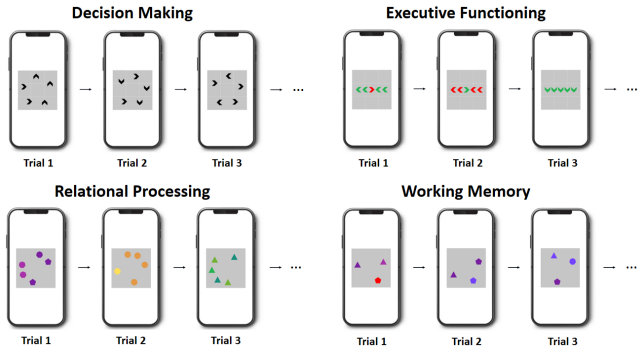

Figure 1: Example tasks for each of the four cognitive domains.

In parallel, each participant underwent five MRI scans at regular intervals, including similar cognitive tasks, to enable comparison between artificial and biological network structures.

Methods

Each participant's data from 12 cognitive-behavioral tasks is used to train an individual recurrent neural network (RNN). The goal is to reproduce response patterns similar to those observed in humans by optimizing model parameters. The RNNs receive encoded input sequences and learn participants' behavior through iterative training.

The networks follow a shallow architecture with three layers: input (77 units), recurrent (256 units), and output (33 units). Layers are fully connected, with linear transformations between layers and nonlinear activation in the recurrent layer. This setup offers a balance between sufficient complexity for task learning and simplicity for subsequent topological analysis. Network parameters are optimized via Backpropagation Through Time (BPTT) using the Adam optimizer. Model selection was done through randomized grid search and with respect to highest performance. After training, the artificial networks are analyzed using graph-theoretical methods to examine topological features.

Results

Performance was measured using a population vector method, counting outputs as correct if within 35° of the participant's response. ANN training accuracy ranged from 0.7 to 1.0, and test accuracy from 0.5 to 1.0, indicating that ANNs can learn to predict individual behavior from the cognitive task battery.

Figure 2 shows results from a stable data phase (months 3–5) with consistent task difficulty and behavior. It displays RNN training (top) and test (bottom) performance, along with hidden layer functional correlation (Gram) matrices used to extract topological markers.

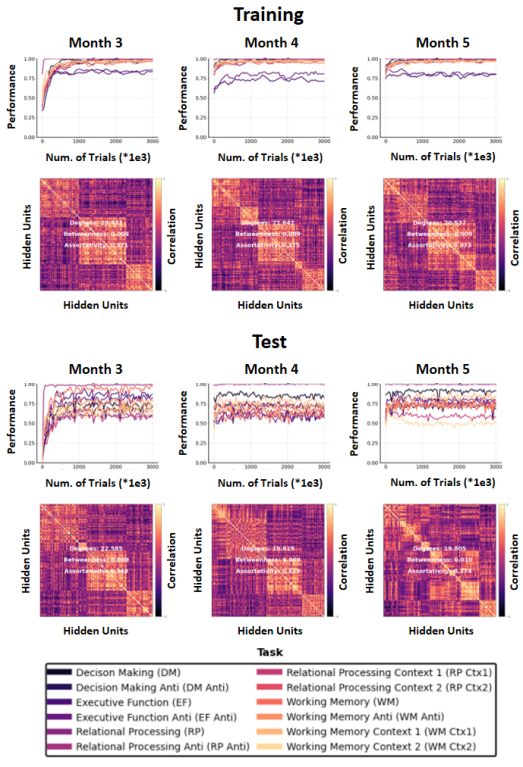

Figure 2: Training and test performance of a model trained on data from a healthy participant for three consecutive months. Below, the corresponding functional correlation matrices of the models' units in the hidden layer are shown, respectively.

To evaluate the within-subject robustness of our results, neural networks were trained multiple times using the same architecture and data, varying only in initial weight initialization and data shuffling. We then analyzed topological markers — such as average degree, average betweenness, and assortativity — both within and across participants.

Results showed that within-month distributions of markers exhibited very low variance (2 out of 3 markers < 0.01). Within-subject marker distributions across months were generally stable, except in cases where task difficulty was adapted (13 out of 24 comparisons (t-test) were non-significant). In contrast, between-subject differences were largely significant (5 out of 6 comparisons (t-test), $p < .05$), with the only exception being the comparison between the healthy and the depressed participant.

These findings indicate that topological network properties are reliable within and meaningfully different between individuals.

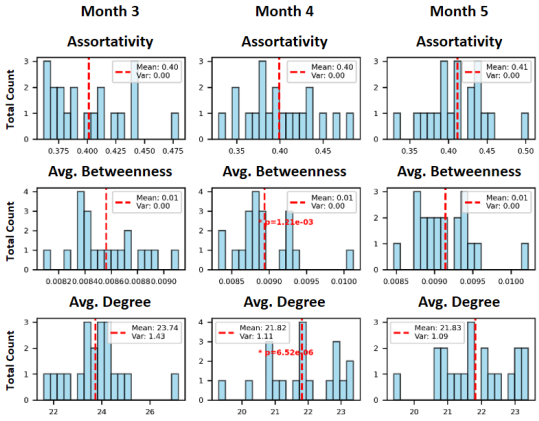

Figure 3: Topological Marker distributions of 20 models trained with the same hyperparameters on healthy control data.

Discussion

The results demonstrate that individually trained RNNs can learn the cognitive behavior of real participants with distinct cognitive profiles. In doing so, the networks capture robust and significant differences between participants in the topology of their hidden layers, while maintaining reliable consistency within participants. During stable phases — characterized by constant task difficulty and participant performance — topological markers remain consistent across training runs. However, when task difficulty or behavior changes, these marker distributions shift significantly.

These findings show that RNNs can distinguish individual cognitive profiles and track changes over time. This approach could help model the dynamics of mental disorders and explore how neural networks align with brain architecture to further study cognition in health and disease.

Code availability

All necessary code for training and analyzing the networks described here is available on GitHub:

`https://github.com/oliver-frank/art_beRNN`

References

Barak, O. (2017). Recurrent neural networks as versatile tools of neuroscience research. *Current Opinion in Neurobiology*, *46*, 1–6. doi: 10.1016/j.conb.2017.06.003

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364. doi: 10.1038/nn.4502

Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T., & Wang, X.-J. (2019). Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience*, *22*(2), 297–306. doi: 10.1038/s41593-018-0310-2