

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

A Principled Framework for Individual Differences in Neural Networks

Permalink

<https://escholarship.org/uc/item/6xq1c74b>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Mathony, Marvin

Modirshanechi, Alireza

Schulz, Eric

Publication Date

2026

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

A Principled Framework for Individual Differences in Neural Networks

Marvin Mathony, Alireza Modirshanechi, Eric Schulz

Computational modeling is a widely employed method in characterizing individual differences in cognitive mechanisms. However, as cognitive science increasingly turns to complex sequence models, like recurrent neural networks (RNNs), it remains an open question whether and how existing methods for modeling individual differences can be applied to these models. Here, we provide a theoretical framework that generalizes and applies the traditional notion of individual differences to complex sequence models. Using this framework, we design a powerful algorithm that enables the characterization of individual differences using RNN models of cognition. Using synthetic and human data, we show that our theory-guided algorithm robustly outperforms existing heuristics in discovering individual differences from RNNs trained via the vanilla training pipeline. Explicitly accounting for individual differences enables our algorithm to achieve a better behavioral fit than vanilla RNNs. Hence, we establish a principled way to discover individual differences in cognitive mechanisms from behavioral data.