Signatures of discrete action symbols emerge in a task-optimized neural model

Carlos G. Correa (cgc392@nyu.edu)[1,*], Lucas Y. Tian (ltian@rockefeller.edu)[2,3,*], Yue Liu[4], Sreejan Kumar[5], Daniel J Hanuska[2], Kedar Garzón Gupta[5], Xiao-Jing Wang[4], Winrich A. Freiwald[2], Joshua B. Tenenbaum[3,6] & Marcelo G. Mattar[1,4]

[1]Department of Psychology, New York University, New York, NY, USA
[2]Laboratory of Neural Systems, The Rockefeller University, New York, NY, USA
[3]Center for Brains, Minds and Machines, Massachusetts Institute of Technology, Cambridge, MA, USA
[4]Center for Neural Science, New York University, New York, NY, USA
[5]Zuckerman Mind, Brain, Behavior Institute, Columbia University, New York, NY, USA
[6]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[*]Equal contribution

Abstract

Key to intelligence is the ability to flexibly compose and combine atomic concepts into representations that guide behavior, making "infinite use of finite means". However, modeling often assumes discrete concepts exist and focuses on their *use*, leaving it unclear how they are *learned* from complex sensory inputs and mapped to motor outputs. To address this limitation, we build on a drawing-like task and behavioral metrics that indicate discrete structure in motor behavior from Tian et al. (2025). Importantly, these discrete concepts were encoded in neural recordings from non-human primates that exhibited these behavioral metrics, validating the metrics. Here, we use this task and behavioral metrics as a modeling target, to quantify how discrete structure arises in a task-optimized neuro-symbolic model (Liang et al., 2022). Our model recapitulates these behavioral metrics, offering a foundation to compare with neural data and understand what drives the emergence of discrete conceptual structure.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*