UCDAVIS
CENTER FOR NEUROSCIENCE

Randall C. O'Reilly

Professor

Departments of Psychology & Computer Science, Center for Neuroscience

1544 Newton Ct   (530) 757-8870
Davis, CA 95816   Email: oreilly@ucdavis.edu

November 7, 2019

Dear Editor:

We are hereby submitting our manuscript entitled "Deep Predictive Learning as a Model for Human Learning" for publication in *Nature*.

This paper presents a significant advance in understanding how the human brain learns, based on the idea that canonical circuits between the neocortex and thalamus drive alternating phases of prediction and bottom-up outcomes, and the resulting prediction errors (as temporal differences) can drive powerful learning. Critically, we show for the first time that learning based solely on predicting raw visual inputs can generate higher-level abstract categorical representations of 3D objects, which previously has required explicit human-labeled training. This captures the seemingly magic way in which human learning can create knowledge out of raw experience, without explicit teaching. Thus, we argue that this work brings us significantly closer to a fully plausible and powerful computational framework for understanding learning in the neocortex, while building on the recent excitement about advances in deep neural network learning algorithms (which we directly evaluate in comparison to our biologically-based model).

Thank you for considering our work.

Sincerely,

Randall O'Reilly