

Neural Foundations of Mental Simulation: Future Prediction of Latent Representations on Dynamic Scenes

Aran Nayebi,¹ Rishi Rajalingham^{1,4}, Mehrdad Jazayeri^{1,2}, Guangyu Robert Yang^{1,2,3}

¹McGovern Institute for Brain Research, MIT; Cambridge, MA 02139

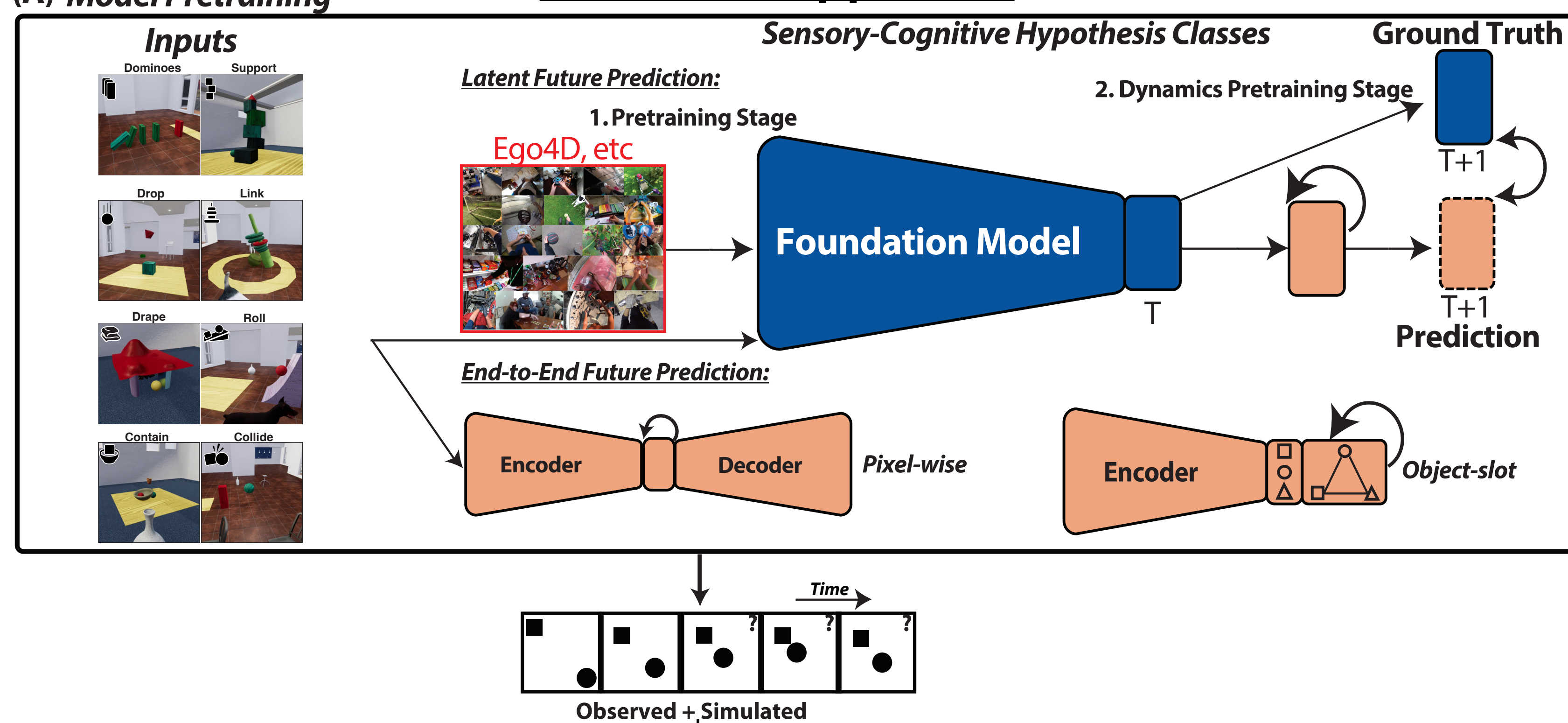
³Department of Electrical Engineering and Computer Science, MIT; Cambridge, MA 02139

²Department of Brain and Cognitive Sciences, MIT; Cambridge, MA 02139

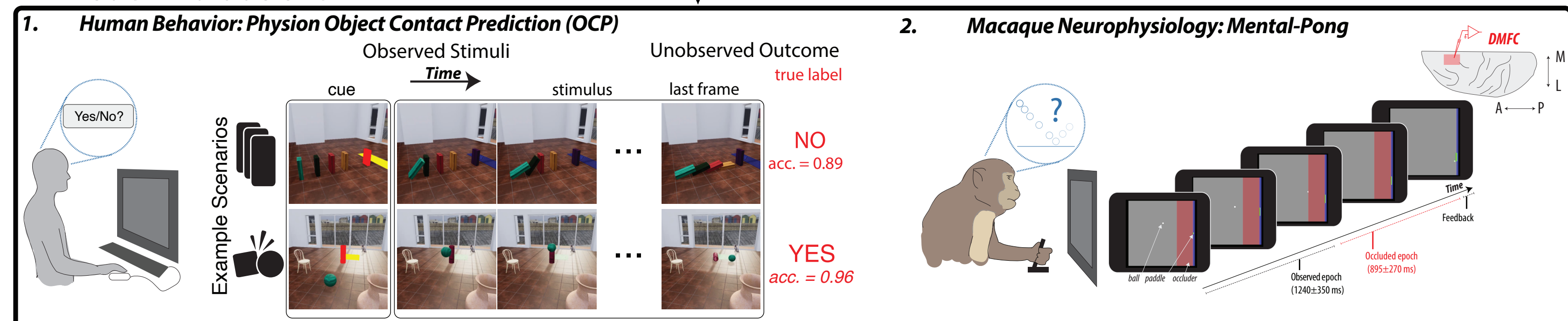
⁴Reality Labs, Meta; 390 9th Ave, New York, NY 10001

(A) Model Pretraining

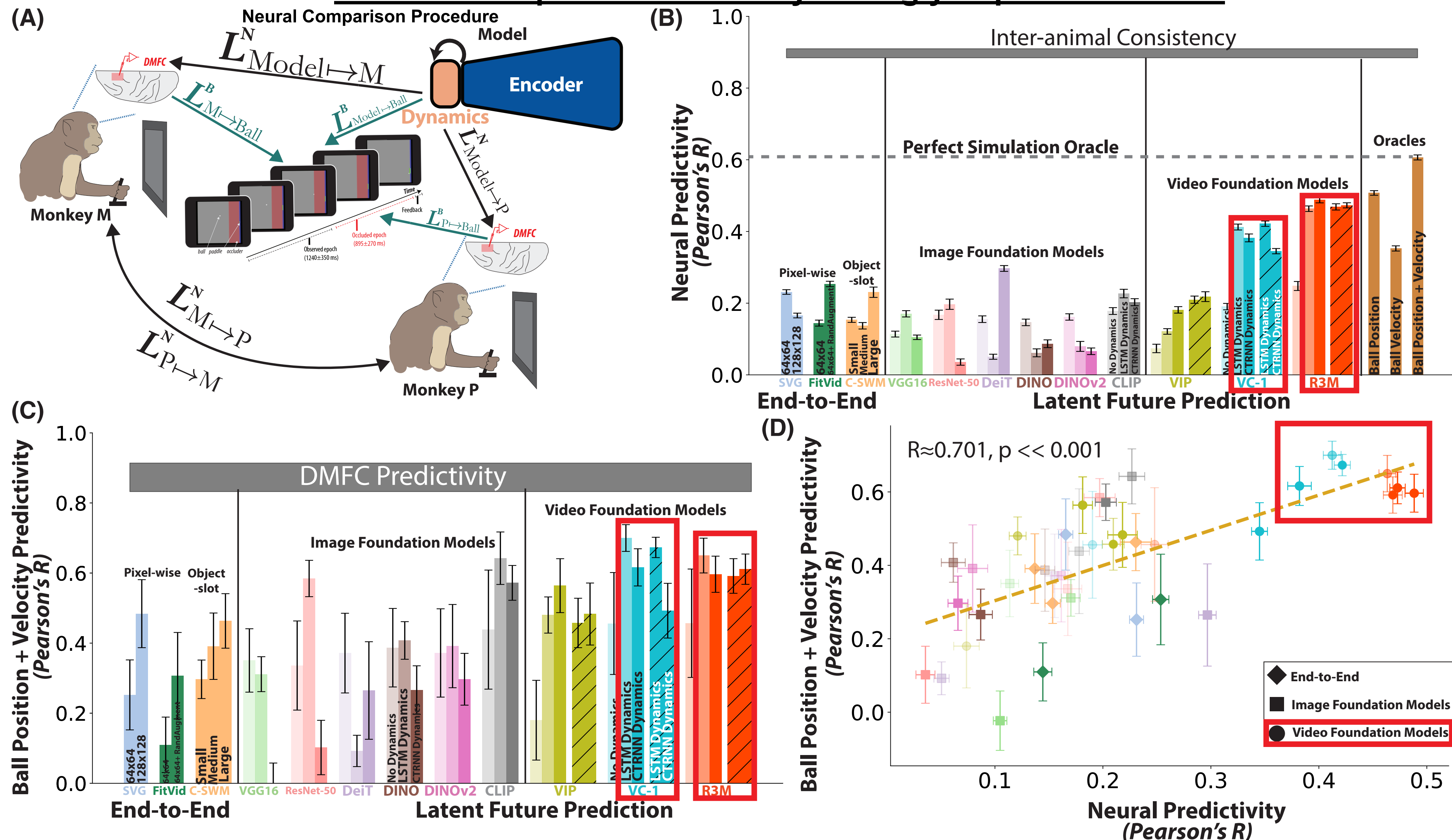
1. Overall Approach



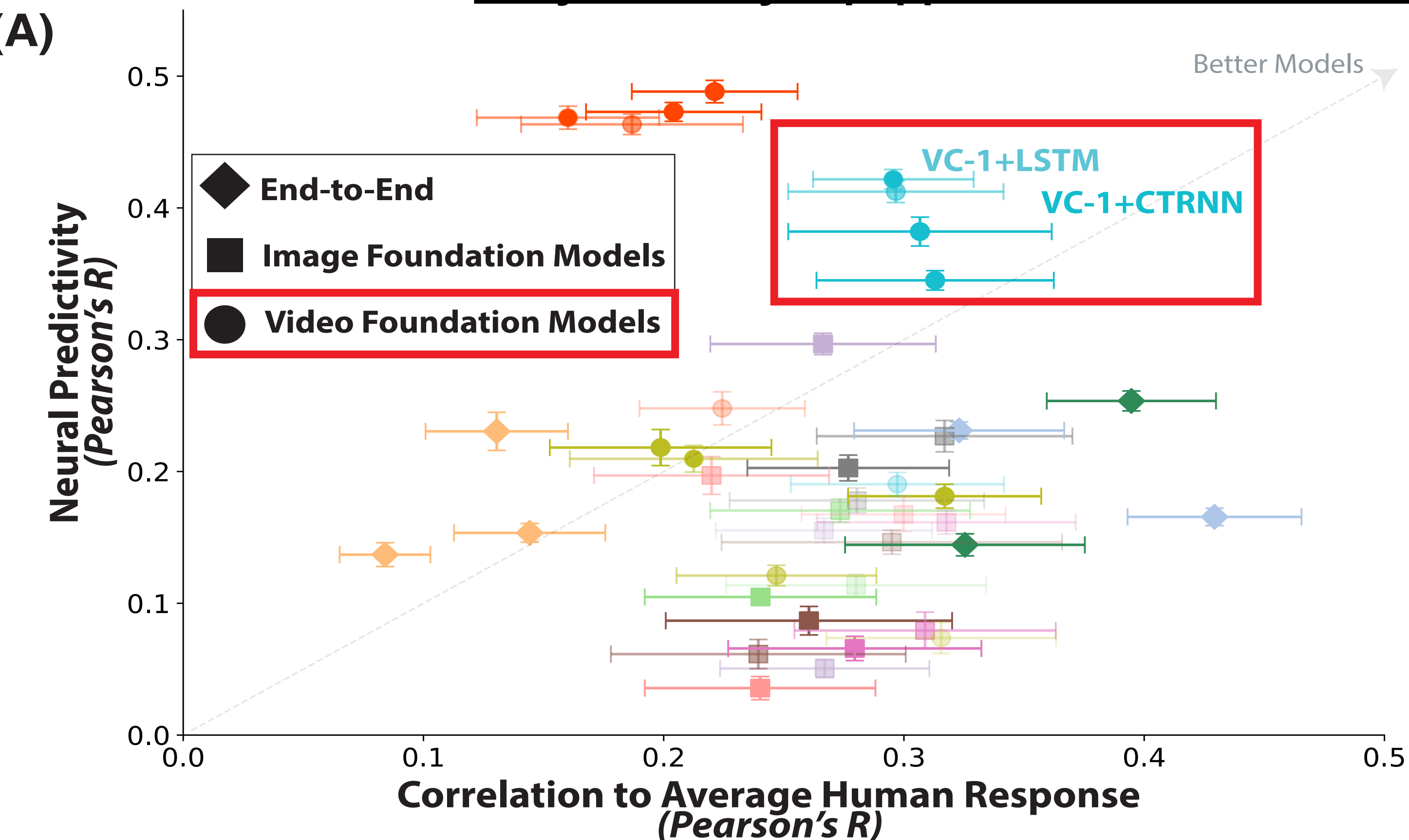
(B) Model Evaluations



2. Neural Response Predictivity Strongly Separates Models



3. Dynamically-Equipped Video Foundation Models Can Match Both Human Behavioral & Neural Response Patterns Across Environments



(B)

