Neural Networks Need Real-World Behavior

Invited Commentary on "Deep Problems with Neural Network Models of Human Vision" (Bowers et al., 2022)

Aedan Y. Li[1] and Marieke Mur[1,2]

[1]Department of Psychology, Western University

[2]Department of Computer Science, Western University

Author Note:

For correspondence, please contact Aedan Y. Li (aedan.li@uwo.ca) and Marieke Mur (mmur@uwo.ca). *Funding Statement:* AYL is supported by a BrainsCAN Postdoctoral Fellowship. MM is supported by an NSERC Discovery Grant. *Conflict of interest*: None declared.

Bowers et al. (2022) propose to use controlled behavioral experiments when evaluating deep neural networks as models of biological vision. We agree with the sentiment and draw parallels to the notion that "neuroscience needs behavior". As a promising path forward, we suggest complementing image recognition tasks with increasingly realistic and well-controlled task environments that engage real-world object recognition behavior.

Keywords: Real-world tasks, Continuous behavioral dynamics, Virtual reality, Deep neural networks, Object recognition

Bowers et al. describe the importance of targeted behavioral experiments when evaluating deep neural networks as models of biological vision. We agree with the sentiment and draw parallels to the notion that "neuroscience needs behavior" (Krakauer et al., 2017). A major point raised by Bowers et al. is that one system – a neural network – can provide an excellent prediction of another system – the visual system – while relying on entirely different mechanisms. Carefully designed behavioral experiments are needed to assess how good the match really is. This point echoes the historic *multiple realizability* argument highlighted by Krakauer et al., which states that different (neural) mechanisms can solve the same computational problem. Krakauer and colleagues proposed the same solution: carefully designed behavioral experiments, to generate and test hypotheses about the neural mechanisms that give rise to behavior. In essence, neuroscience and modeling both need behavior to guide hypothesis testing and theory development in their endeavor to understand how the brain works.

What types of behavioral experiments are best suited to evaluate deep neural networks as models of biological vision? As suggestions for the modeling community, we take inspiration from solutions pioneered by neuroscience in recent years (e.g., Snow & Culham, 2021). There is growing realization that real-world object recognition engages distinct neural responses compared to the behaviors involved with standard image recognition tasks. In the traditional experiment, observers respond with button presses to images displayed on a computer monitor as brain activity is recorded. This approach has provided important insights on biological vision and has served as a great starting point for model evaluation (e.g., Jozwik et al. 2023). However, traditional experiments do not fully capture how humans interact with objects in real-world environments.

We suggest that our experiments should increasingly mimic real-world behavior, by: 1) including tasks beyond image recognition when evaluating deep neural networks, and 2) developing platforms that enable simulation of realistic task environments. Using these environments, both humans and models can be subjected to a wide range of real-world behavioral tasks such as object tracking (e.g., following a moving animal) or visual search (e.g., finding objects in cluttered scenes); also see Peters & Kriegeskorte (2021) for discussions. The researcher will be offered a level of control that supports carefully designed experiments while maintaining ecological validity. The proposed platforms are now within reach thanks to advances in virtual reality and 3D computer graphics, which are yielding powerful game engines accessible to psychologists and modelers alike. Promising recent approaches have extended the *Unity* game engine to the design of psychology experiments (e.g., Alsbury-Nealy et al., 2022; Brookes et al., 2020; Peters et al., 2022; Starrett et al., 2021) and the simulation of interactive physics (e.g., *ThreeDWorld;* Gan et al., 2021).

Importantly, we suggest that the behavior in task environments should include the measurement of continuous dependent variables that unfold over time. Traditional cognitive psychology and neuroscience experiments use binary metrics such as "yes/no" or "multiple-choice" questions with one correct option among competitors (e.g., image classification). By contrast, humans in the real world have evolved to complete unstructured tasks in service of survival-related goals. We use cognitive abilities honed through millions of years of primate evolution and over a decade of childhood development to navigate environments, build tools, find food, solve problems, and interact with other humans in cooperative and competitive settings. These dynamic behaviors involve head, body, and limb movements (Adolph & Franchak, 2017) and are based on internal decisions made from the input received from our sensory organs at millisecond timescales (Stanford et al., 2010). Measuring the continuous behavioral dynamics may allow for richer understanding compared to discrete variables that average over many experimental trials (Spivey, 2007; for object memory dynamics, see Li et al., 2023; for navigation dynamics, see de Cothi et al., 2022; for "continuous psychophysics", see Straub & Rothkopf, 2022).

The models we build should also explain neural activity measured as humans complete different experimental tasks. Not only will this approach create a wealth of interdisciplinary opportunities, but modelers could take advantage of psychology and neuroscience theory which continues to make important predictions about behavior (e.g., Behrens et al., 2018; Cowell, Barense, & Sadil, 2019). As one example, the anterior temporal lobes are theorized to be a centralized "hub" region of the human brain involved in combining multiple sensory features to form object concepts (Lambon Ralph et al., 2017). This structure supports the formation of new concepts in tasks involving the combination of 3-dimensional shape and sound (Li et al., 2022). Furthermore, damage to the anterior temporal lobes results in predictable impairments on memory, perception, and learning tasks (i.e., *semantic dementia*; Hodges & Patterson, 2007; Barense et al., 2010). A complete model should be able to make novel predictions about behavioral and brain responses while also accounting for existing data across many tasks.

We have outlined concrete suggestions toward a collaborative path that we envision to be productive. We suggest that modelers should design realistic tasks in virtual reality, measure the continuous behavioral dynamics that unfold over time, and assess correspondences to brain activity across many tasks. However, there are also many challenges that lie ahead before these suggestions can be fully realized: the expertise required to span cognitive psychology and neuroscience in addition to computational modeling is daunting. Developing naturalistic real-world experiments requires programming skills often not taught in psychology and neuroscience curriculums, whereas theoretical models important for understanding human cognition are often not taught in computer science. Fully characterizing the dynamics of behavior and brain activity will likely require theory and measurement techniques that have not yet been developed (Druckmann & Rust, 2023). For these reasons, we suggest an incremental, highly interdisciplinary and collaborative approach toward real-world experiments, which we hope will lead to a more complete understanding of how the human brain may support object-centered representations.

Our suggestions reemphasize the centrality of behavior – described as "psychological findings" by Bowers et. al – across both the development of more human-like neural networks as well as in the continued understanding of the human brain.

References

Adolph, K. E., & Franchak, J. M. (2017). The development of motor behavior. *Wiley Interdiscip Rev Cogn Sci.*, *8*(1-2), 10.1002/wcs.1430. doi: 10.1002/wcs.1430.

Alsbury-Nealy, K., Wang, H., Howarth, C. *et al.* (2022). OpenMaze: An open-source toolbox for creating virtual navigation experiments. *Behav Res*, *54*, 1374–1387. https://doi.org/10.3758/s13428-021-01664-9

Barense, M. D., Rogers, T. T., Bussey, T. J., Saksida, L. M., & Graham, K. S. (2010). Influence of conceptual knowledge on visual object discrimination: insights from semantic dementia and MTL amnesia. *Cerebral Cortex, 20*(11), 2568–2582. https://doi.org/10.1093/cercor/bhq004

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, *100*(2), 490–509. https://doi.org/10.1016/j.neuron.2018.10.002

Brookes, J., Warburton, M., Alghadier, M. *et al.* (2020). Studying human behavior with virtual reality: The Unity Experiment Framework. *Behav Res*, *52*, 455–463 (2020). https://doi.org/10.3758/s13428-019-01242-0

Bowers, J. S., Malhotra, G., Dujmović, M., Montero, M. L., Tsvetkov, C., Biscione, V., Puebla, G., Adolfi, F., Hummel, J. E., Heaton, R. F., Evans, B. D., Mitchell, J., & Blything, R. (2022). Deep Problems with Neural Network Models of Human Vision. *The Behavioral and brain sciences*, 1–74. Advance online publication. https://doi.org/10.1017/S0140525X22002813

Cowell, R. A., Barense, M. D., & Sadil, P. S. (2019). A Roadmap for Understanding Memory: Decomposing Cognitive Processes into Operations and Representations. *eNeuro*, *6*(4), ENEURO.0122-19.2019. https://doi.org/10.1523/ENEURO.0122-19.2019

de Cothi, W., Nyberg, N., Griesbauer, E. M., Ghanamé, C., Zisch, F., Lefort, J. M., Fletcher, L., Newton, C., Renaudineau, S., Bendor, D., Grieves, R., Duvelle, É., Barry, C., & Spiers, H. J. (2022). Predictive maps in rats and humans for spatial navigation. *Current Biology: CB*, 32(17), 3676–3689.e5. https://doi.org/10.1016/j.cub.2022.06.090

Druckmann, S., & Rust, N. C. (2023). Unraveling the entangled brain: How do we go about it? *Journal of Cognitive Neuroscience*, *35*, 368–371. https://doi.org/10.1162/jocn_a_01950

Gan, C., Schwartz, J., Alter, S. e*t al.*, 2021. ThreeDWorld: A platform for interactive multi-modal physical simulation. *bioRxiv*. https://doi.org/10.48550/arXiv.2007.04954

Hodges, J. R., & Patterson, K. (2007). Semantic dementia: a unique clinicopathological syndrome. *The Lancet. Neurology*, *6*(11), 1004–1014. https://doi.org/10.1016/S1474-4422(07)70266-1

Jozwik, K.M., Kietzmann, T.C., Cichy, R.M., Kriegeskorte, N., & Mur, M. (2023). Deep neural networks and visuo-semantic models explain complementary components of human ventral-stream representational dynamics. *The Journal of Neuroscience.* https://doi.org/10.1523/JNEUROSCI.1424-22.2022

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, *93*(3), 480–490. https://doi.org/10.1016/j.neuron.2016.12.041

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience, 18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Li, A.Y., Yuan, J.Y., Pun, C., & Barense, M. D. (2023). The effect of memory load on object reconstruction: Insights from an online mouse-tracking task. *Atten Percept Psychophys*. https://doi.org/10.3758/s13414-022-02650-9

Li, A.Y., Ladyka-Wojcik, N., Qazilbash, H., et al. (2022). Multimodal object representations rely on integrative coding. *bioRxiv*. https://doi.org/10.1101/2022.08.31.504599

Peters, B., & Kriegeskorte, N. (2021). Capturing the objects of vision with neural networks. *Nature Human Behaviour*, *5*(9), 1127–1144. https://doi.org/10.1038/s41562-021-01194-6

Peters, B., Retchin, M., & Kriegeskorte, N. (2022). Flying Objects: Challenging humans and machines in dynamic object vision. *Cognitive Computational Neuroscience*. https://doi.org/10.32470/ccn.2022.1301-0

Snow, J. C., & Culham, J. C. (2021). The Treachery of Images: How Realism Influences Brain and Behavior. *Trends in cognitive sciences*, *25*(6), 506–519. https://doi.org/10.1016/j.tics.2021.02.008

Spivey, M. (2007). *The Continuity of Mind.* Oxford University Press.

Stanford, T. R., Shankar, S., Massoglia, D. P., Costello, M. G., & Salinas, E. (2010). Perceptual decision making in less than 30 milliseconds. *Nature Neuroscience*, *13*(3), 379–385. https://doi.org/10.1038/nn.2485

Starrett, M.J., McAvan, A.S., Huffman, D.J. *et al.* (2021). Landmarks: A solution for spatial navigation and memory experiments in virtual reality. *Behav Res*, *53*, 1046–1059 (2021). https://doi.org/10.3758/s13428-020-01481-6

Straub, D., & Rothkopf, C. A. (2022). Putting perception into action with inverse optimal control for continuous psychophysics. *eLife*, *11*, e76635. https://doi.org/10.7554/eLife.76635