

Beyond task-optimized neural models: constraints from embodied cognition

Kaushik Lakshminarasimhan (jl5649@columbia.edu)

Columbia University, 3227 Broadway, New York, NY 10027 USA

Akis Stavropoulos (ges6@nyu.edu)

New York University, 4 Washington Place, New York, NY 10004 USA

Dora Angelaki (da93@nyu.edu)

New York University, 4 Washington Place, New York, NY 10004 USA

Abstract

Generic neural networks optimized for task performance are often very successful in predicting neural activity in animals. However, neural mechanisms dictate not only the task performance but also *how* a particular task is solved. Can we deduce mechanisms from cognitive strategies? To find out, we asked humans and monkeys to perform a challenging task in which they steered to a remembered goal location by integrating self-motion in a virtual environment lacking position cues. Although this task requires only mentally tracking one's position relative to the goal, participants physically tracked this latent task variable with their gaze – an instance of embodied cognition. Restraining eye movements worsened task performance suggesting that embodiment plays a computational role. Above findings are well explained by a neural model with tuned bidirectional connections between oculomotor circuits and circuits that integrate sensory input. In contrast to other task-optimized models, this model correctly predicted that leading principal components of the monkey posterior parietal cortex activity should encode their position relative to the goal. These results explain the computational significance of motor signals in evidence-integrating circuits and suggest that plasticity between those circuits might enable efficient learning of complex tasks via embodied cognition.

Keywords: recurrent neural network; embodied cognition; neural representation; posterior parietal cortex; navigation

Introduction

Neural network models trained to perform neuroscience tasks are generally good at also predicting the response properties of brain areas that drive behavior in those tasks (Yamins & DiCarlo, 2016; Yang, Joglekar, Song, Newsome, & Wang, 2019). However, such models are typically grounded in generic neural architectures – feedforward or recurrent depending on the task – and cannot explain why neural computations are distributed across functionally distinct circuits. A jarring example of distributed brain computation is the prevalence of motor signals in sensory and association areas (Musall, Kaufman, Juavinett, Gluf, & Churchland, 2019). To understand such phenomena, we need to additionally factor in the *cognitive strategy* used by animals to solve the task. Common neuroscience tasks like binary decision-making are too simple to admit interesting cognitive strategies, especially when

participants are mechanically restrained. Instead, we used a challenging navigation task featuring action-perception loops in virtual reality (VR) and free eye movements. We analyzed their behavioral strategy and used this to constrain the space of neural models. The resulting model explained neural data in monkeys better than alternative models that were optimized for the task but not constrained by behavioral strategy.

Results

Humans and monkeys performed a navigation task in which they manipulated a handheld joystick to steer to a transiently cued goal location in a VR environment without explicit position cues (i.e., stable landmarks) (Figure 1A). At the beginning of each trial, a circular goal blinked briefly (~300ms) at a random location within the field of view on the ground plane, and then disappeared. The joystick controlled forward and angular velocities, allowing free movement in 2D (Figure 1B). The task was to steer and stop as close to the goal as possible.

Behavior

To perform the task correctly, participants must deduce their movement velocity at each moment – from optic flow (visual condition) or inertial cues (vestibular condition) – and integrate it over time to keep track of their position relative to the goal. They were able to successfully perform this in both conditions (Figure 1C) but more remarkably, the above mental computation manifested itself physically in eye movement dynamics: participants kept looking at the (invisible) goal location until they reached it (Figure 1D,E). Critically, restraining gaze by forcing participants to fixate at a location on the screen impaired navigation (Figure 1F). These results extend past findings from visual domain (Lakshminarasimhan et al., 2020) and suggest that participants integrate movement velocity to track their position relative to the goal with two important cognitive constraints: (i) the evolving belief about their position is embodied in eye position, and (ii) this embodiment has a computational role. We now turn to modeling in order to understand the underlying neural mechanisms.

Model

Neurally, this task can be implemented in a recurrent neural network that integrates velocity input to drive appropriate joystick movements depending on the target location. This integration process is thought to take place in the mammalian posterior parietal cortex (PPC). Constraint (i) above implies



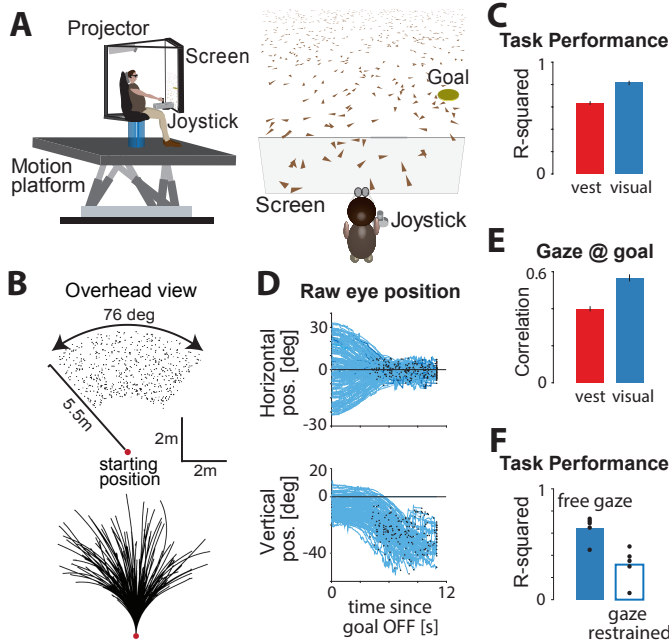


Figure 1: A. Experimental set-up. B. Goal locations (top) and trajectories (bottom) across trials. C. Eye position dynamics – each line is one trial. D. Navigation performance (r-squared, $N=8$ humans). E. Correlation between gaze position and relative goal position. F. Navigation performance under free vs restrained eye movements. Results similar in monkeys ($N=3$).

that PPC communicates participant's position to the oculomotor (OC) network, while constraint (ii) implies that OC activity contributes to task performance. We train different neural models optimized to do this task both with and without the cognitive strategy constraints, and evaluate how well each model predicts behavioral and neural data recorded in monkeys.

We consider four models that are architecturally identical but differ in which connections are tuned (Figure 2A – green). They feature a PPC module comprising 100 recurrently connected nonlinear ('tanh') units that receive sensory inputs, and send projections to motor units that drive joystick actions. The PPC module has bidirectional connections with OC units that drive eye movements. Two of the models are optimized solely for task performance (joystick movements) by tuning either just the readout weights onto motor units (Model 1) or both readout and recurrent weights (Model 2). The two remaining models are optimized for task performance by tuning the readout weights, while also being constrained by the strategy used by humans and monkeys. To satisfy constraint (i), we tune the weights from PPC to OC to minimize an auxiliary loss such that OC could dynamically decode position relative to the target from PPC activity (Models 3 and 4). To satisfy constraint (ii), we additionally tune the feedback projection from OC to PPC to optimize for task performance (Model 4). For simplicity, we ignore recurrence within modules other than PPC. In all cases, weights are tuned by backpropagation through time.

Model performance By construction, only the models constrained by cognitive strategy explain participants' eye movements (Figure 2B). Of these two models, only the one with tuned feedback from OC to PPC had good task performance (Figure 2C). Notably, the performance of this model (Model 4) was almost as good as the task-optimized model in which all recurrent weights are tuned (Model 2) despite having substantially fewer tunable parameters (600 vs 10,200). This suggests that PPC-OC interactions serve as a useful anatomical motif to support task performance: by selectively funneling the task-relevant subspace of PPC activity into OC, Model 4 enables learning efficiently (with fewer resources) thereby highlighting the computational significance of embodied cognition.

Model predictions Unlike other models, Model 4 also recapitulates salient aspects of behavior and neural data recorded in PPC. First, it explains why significant trial-by-trial correlations arise between eye movements and navigation accuracy (Figure 2D). In this model, error in estimating position results in poor goal-tracking by eye, which propagates to joystick movements via tuned feedback connections from OC to PPC. Second, it explains why position signals are found in the *high-variance* modes of monkey PPC activity (Figure 2E). The model compresses high-dimensional PPC activity into a low-dimensional OC activity, which is then fed back to PPC. Therefore, position signals communicated from PPC to OC undergo recurrent amplification and end up dominating PPC activity, rather than being buried in low-variance modes.

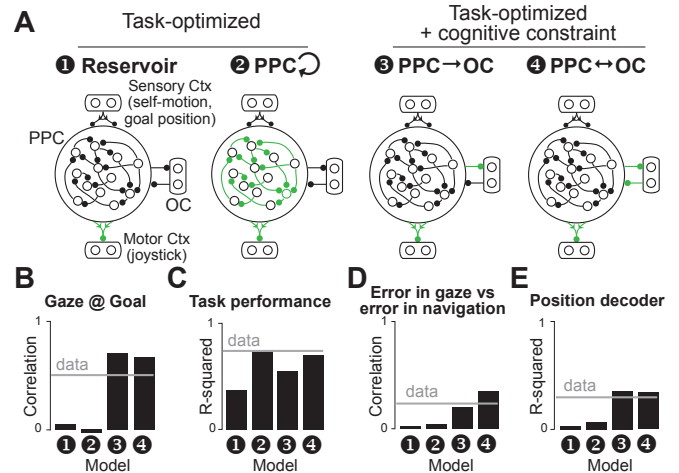


Figure 2: A. Models: tunable weights in green. B. Corr. between gaze position and goal position. C. Navigation performance. D. Corr. between gaze error & navigation error. E. Linear decoding of relative goal position from subspace spanned by 5 leading principal components of PPC activity.

Conclusion

Using a rich behavioral paradigm, we find that a neural model constrained by the cognitive strategy adopted by animals explains neural data better than purely task-optimized models. We propose that mixing of signals between association and

motor areas results from a distributed brain architecture which evolved to implement cognitive computations by grounding subjective beliefs about latent world states in states of the body. Our results are consistent with the *replacement hypothesis* – the idea that our body does most of the work, replacing the need for complex mental representations (Shapiro, 2010). Furthermore, the learning efficiency of the distributed architecture has important implications for realizing artificial general intelligence in robots.

Acknowledgments

This work was supported by the Simons Collaboration on the Global Brain, grant 324143, NIH DC007620, NSF NeuroNex Award DBI-1707398 and the Gatsby Charitable Foundation.

References

- Lakshminarasimhan, K. J., Avila, E., Neyhart, E., DeAngelis, G. C., Pitkow, X., & Angelaki, D. E. (2020). Tracking the mind's eye: Primate gaze behavior during virtual visuomotor navigation reflects belief dynamics. *Neuron*, 106, 662-674. doi: 10.1016/j.neuron.2020.02.023
- Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S., & Churchland, A. K. (2019). Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience*, 22. doi: 10.1038/s41593-019-0502-4
- Shapiro, L. (2010). *Embodied cognition*. London: Routledge.
- Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19. doi: 10.1038/nn.4244
- Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T., & Wang, X. J. (2019). Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience*, 22. doi: 10.1038/s41593-018-0310-2