Topographic ANNs Predict the Behavioral Effects of Causal Perturbations in Primate Visual Ventral Stream IT

Martin Schrimpf, Paul Mc Grath, James DiCarlo

CENTER FOR Brains Minds+ Machines · McGOVERN INSTITUTE · brain+cognitive sciences · MIT

Abstract

Background

The primate visual ventral stream is thought to underlie our ability to recognize objects in the visual world. A series of cortical regions V1, V2, V4, IT transform light hitting our retina into representations that downstream areas read out of.

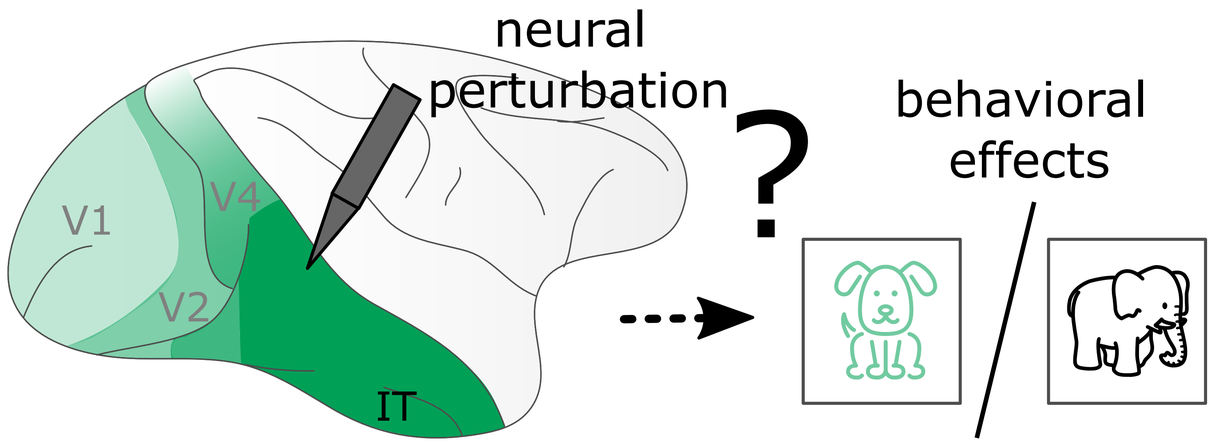
Particular artificial neural network models predict image-elicited neural activity and behavioral outcomes to a first extent (see Brain-Score.org)

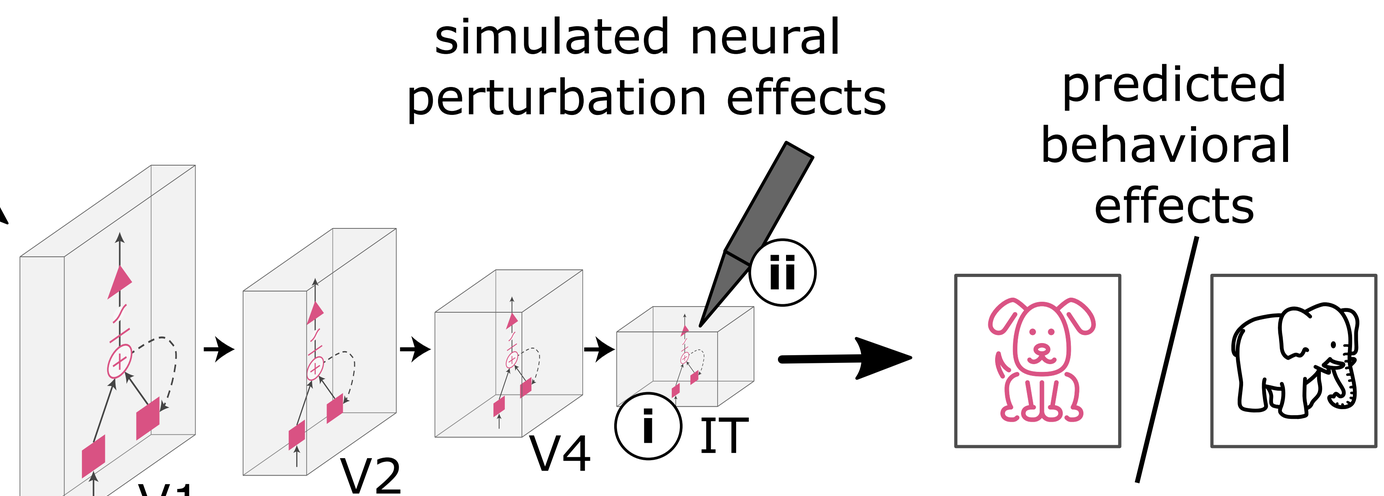
How causal perturbations to neural activity lead to changes in downstream behavior however remains unclear.

Goal in this work

Test causal perturbation experiments on computational models of the primate visual ventral stream and evaluate how closely the model-predicted behaviors match those observed experimentally.

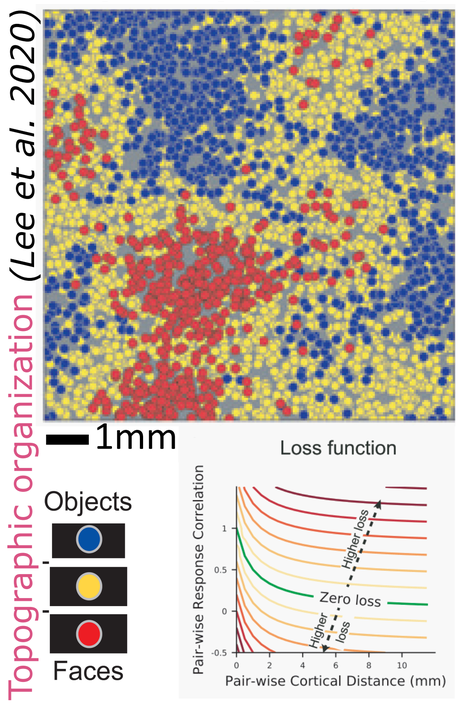
Such predictive models could take us closer to future brain-machine-interfaces.





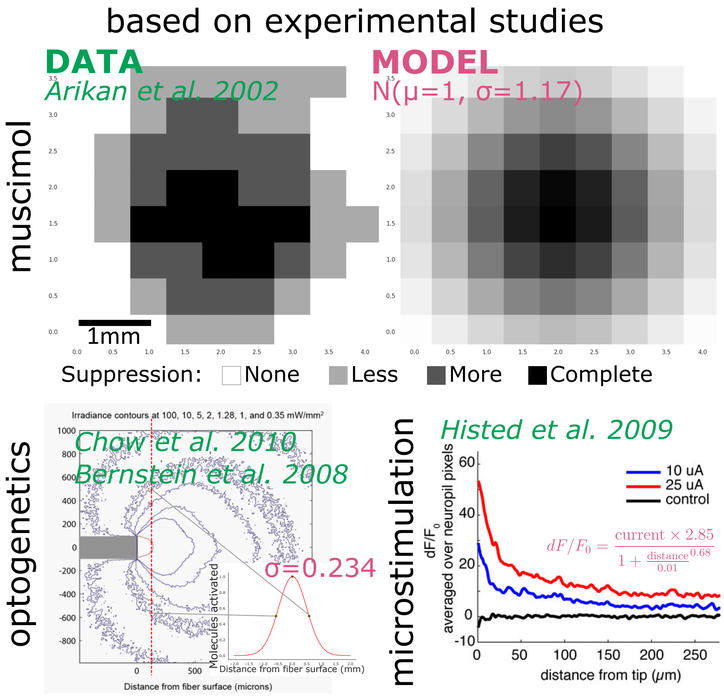
1. Building a "causal" model

(i) Organize model neurons spatially (in tissue)



(ii) Change neural activity following a perturbation

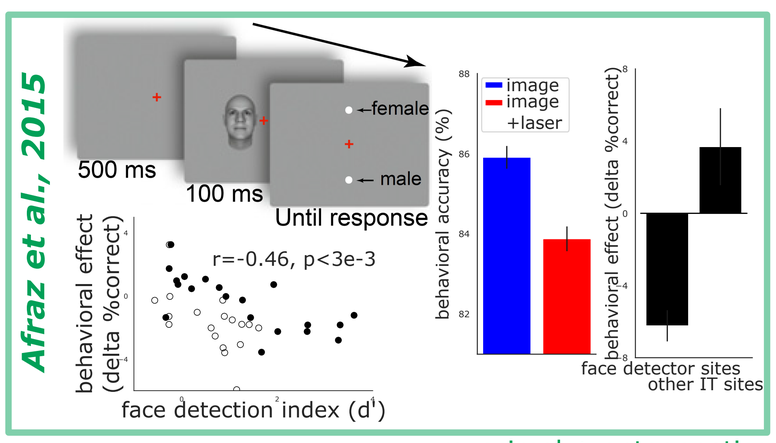
lock down perturbation parameters based on experimental studies



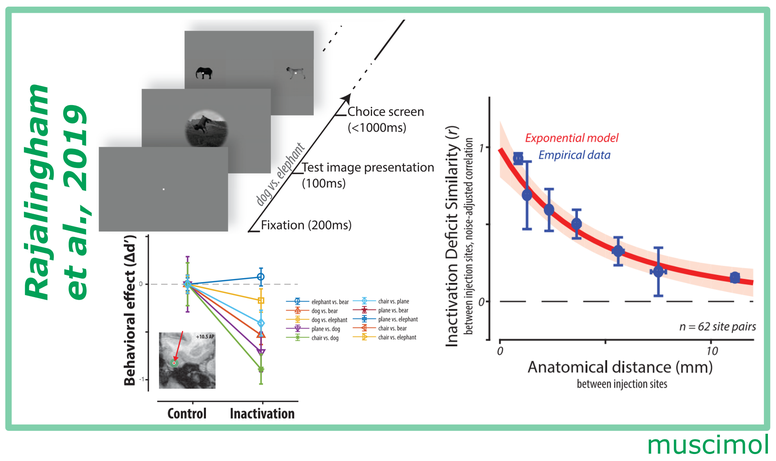
Enriching leading ANN ventral stream models with spatial layout (1) and a perturbation module (2) lets us predict the behavioral effects of neural perturbations in an image-driven task.

2. The combined causal model qualitatively matches a range of experimental data

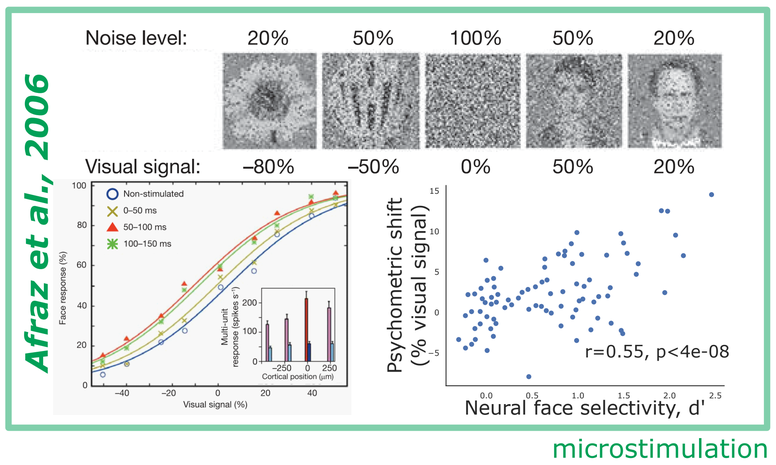
Neural perturbation results



muscimol + optogenetics

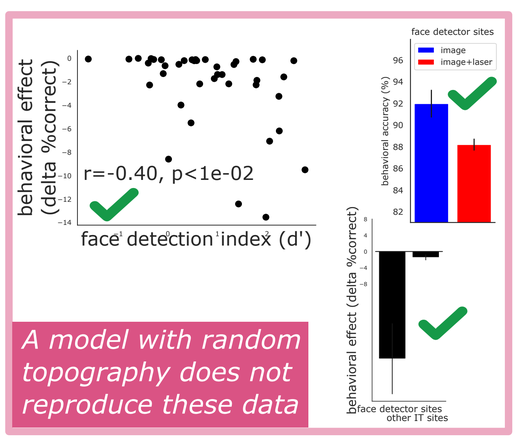


muscimol

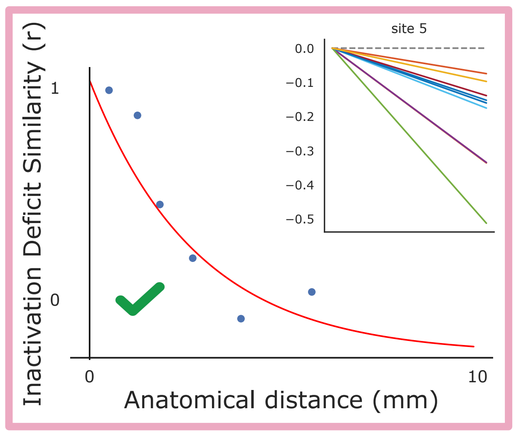


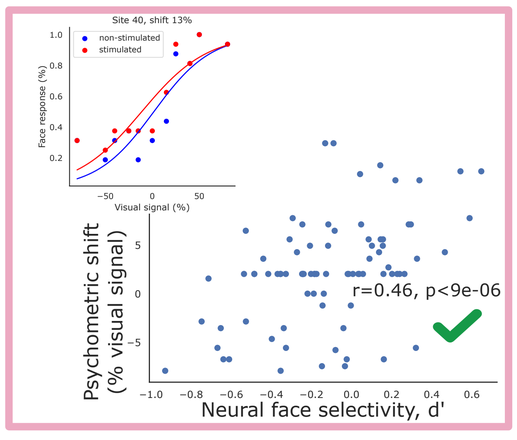
microstimulation

Model predictions



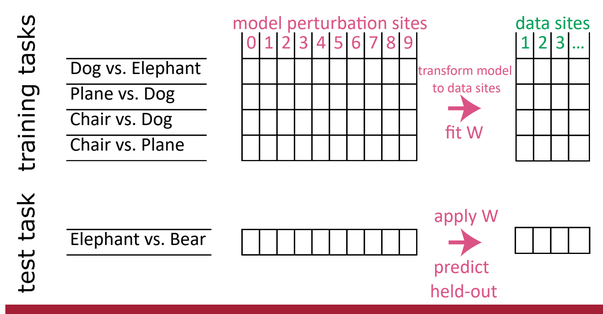
A model with random topography does not reproduce these data

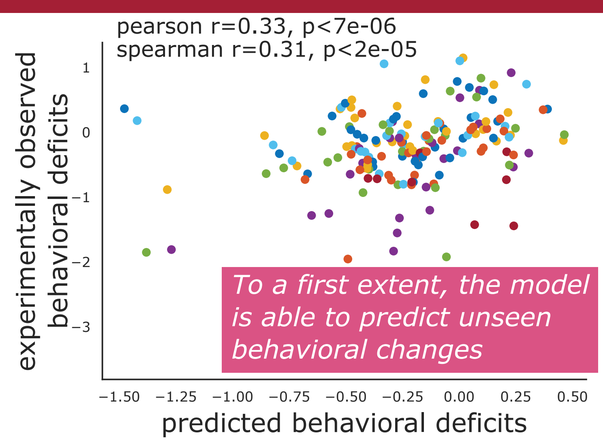




3. Model predicts per-site behavior

We tested model ability to predict precise behavioral changes within Rajalingham2019 injection sites





To a first extent, the model is able to predict unseen behavioral changes

Conclusion

* We found a causal model that qualitatively predicts a range of experimental behavioral observations following neural perturbations
* Quantitative differences exist however - unclear if due to model / data
* We also lack experimental results to lock down magnitude effects of optogenetics and microstimulation on neural activity
* Going forward, we plan model-guided stimulation experiments with predicted stimulation patterns for neural and behavioral desiderata

Funding from Takeda (M.S.), McGovern (M.S.), and the Semiconductor Research Corporation (M.S., J.D.)