Generalization Gradients in Deep Vision Models: Insights from Shepard's Universal Law

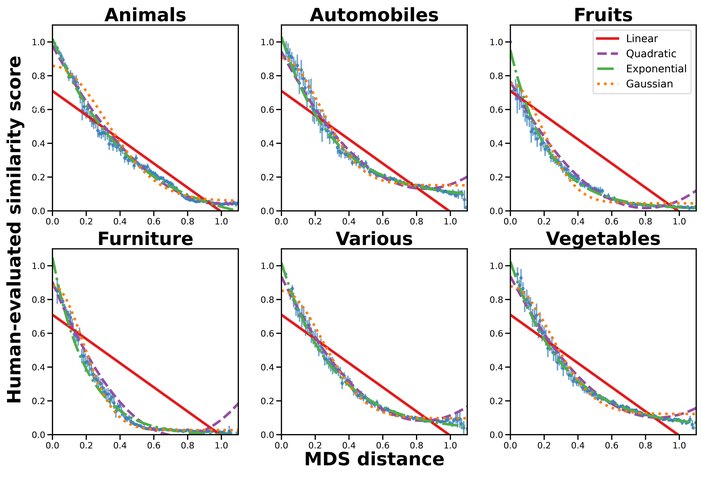
Daniel L. Carstensen[1], Steven M. Frankland[2†], Serra E. Favila[1†]

[1] Brown University, [2] Dartmouth College, [†]Equal Contribution

BROWN DARTMOUTH

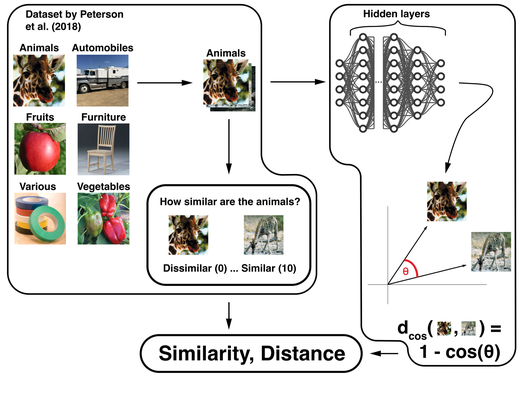
Introduction

- Deep vision models share striking similarities with (and differences to) the primate visual system.
- Shepard's universal law of generalization provides a framework to assess representational alignment.
- **Shepard's law posits that generalization decreases as a concave function of psychological distance reflecting universal cognitive principles.**[1]
- Supported across species and stimuli, and recently validated in large-scale naturalistic image dataset.[2,3]



Does the embedding space of deep vision models align with human psychological similarity space?

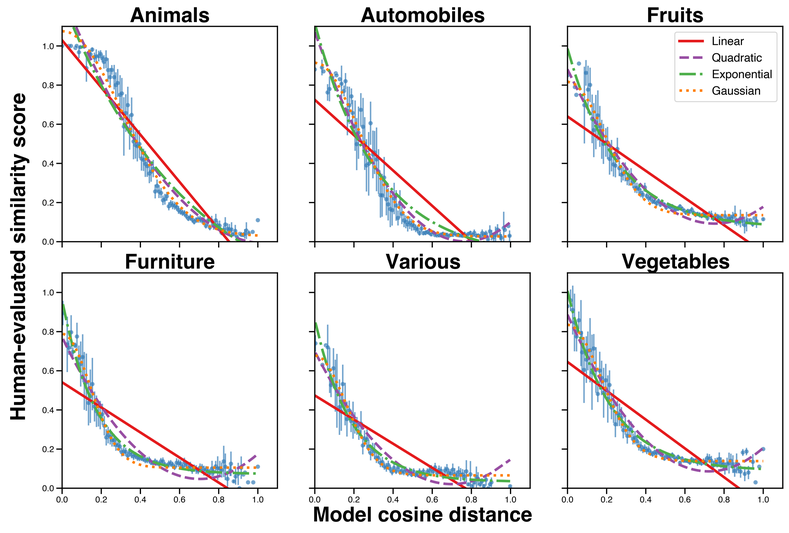
Approach

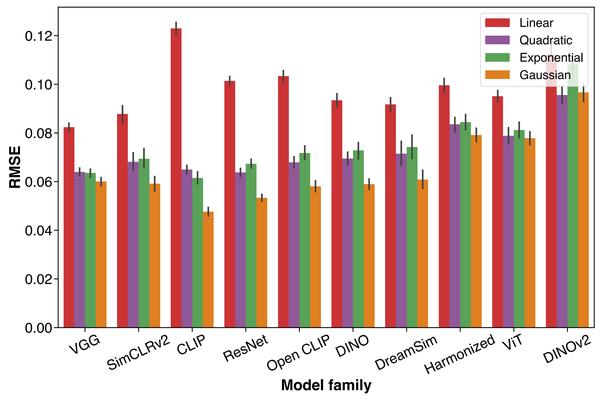


We computed pairwise distances between image embeddings, matched with corresponding human-evaluated similarity scores.

Evidence of Shepard's law

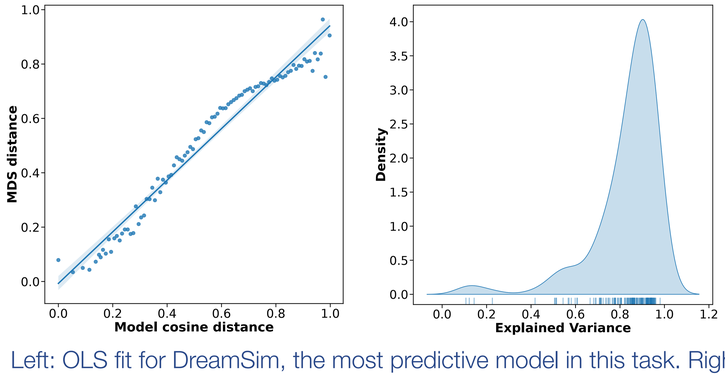
Embedding distances exhibited concave generalization gradients aligned with human similarity judgments.





Top: Generalization gradients for CLIP (ViT-B), the overall best performing model. Left: Average RMSE for curve fits across model families.

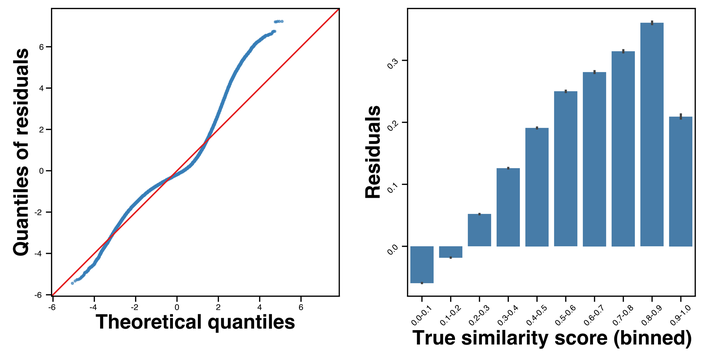
Embedding distances approximated MDS distances well.



Left: OLS fit for DreamSim, the most predictive model in this task. Right: KDE of explained variances across all models and datasets.

Where models deviate

All models systematically overestimated the distance of image pairs with higher similarity scores.



Gaussian-fit residuals pooled across all models and datasets. Left: Normal QQ-plot of residuals. Right: Averaged residuals binned by true similarity ratings (x-axis).

Conclusions & Future directions

Deep vision models often align with Shepard's law, reflecting some alignment with human visual cognition.

Potential next steps:

- Explore the influence of semantic vs. perceptual information on Shepard's law. Preliminary data shows that only later layers of deep vision models exhibit concave gradients, suggesting that they are not driven by low-level visual information.
- Examine possible connections between fMRI image representations, model embedding spaces, and human similarity judgements within the framework of Shepard's law.
- Determine necessary and sufficient conditions of Shepard's law by training custom vision models.

References & Contact

1. Shepard, R. N. (1987). Science
2. Marjieh, R., Jacoby, N., Peterson, J. C., & Griffiths, T. L. (2024). J Exp Psychol Gen
3. Peterson, J. C., Abbott, J. T., & Griffiths, T. L. (2018). Cogn Sci

Contact: daniel_carstensen@brown.edu