Assessing the similarity of cortical object and scene representations through cross-validated voxel encoding models

Nicholas M. Blauch[1,*], Filipe De Avila Belbute Peres[2], Juhi Farooqui[1], Alireza Chaman Zar[3], David C. Plaut[1,4], Marlene Behrmann[1,4]

[1]Center for the Neural Basis of Cognition, Carnegie Mellon Univ., [2]Computer Science Dept., Carnegie Mellon Univ.,
[3]Dept. of Electrical and Computer Engineering, Carnegie Mellon Univ., [1] [4]Dept. of Psychology, Carnegie Mellon Univ.

*blauch@cmu.edu

Introduction

- To what extent are brain mechanisms subserving object and scene recognition distinct?
- Functional imaging reveals reliable response selectivity for scenes in 3 cortical areas in parahippocampal, occipital, and medial parietal cortex (PPA, OPA, and RSC/MPA), and object-selectivity in lateral occipital cortex (LOC)
- Despite its ability to reveal large-scale organizational principles, mean selectivity cannot conclusively rule about single/distinct underlying mechanisms, and is uninformative about the details of representation
- Representational Similarity Analysis[4] provides detailed information about representations but typically requires a common set of stimuli across regions/models.
- To explore the nature of mechanism/s underlying object and scene representations in these regions and throughout cortex, we develop an approach using cross-validated voxel encoding models
- By training a voxel encoding model on one set of stimuli (e.g. objects in ImageNet), we cross-validate that model with prediction on a different stimulus set (e.g. scene images from SUN categories)
- Cross-database generalization serves as a metric of representational similarity/mechanistic overlap of distinct stimuli

Method

- Use BOLD5000: ~5000 images during fMRI for 3 subjects
- Voxel activity was extracted using single-trial GLMs with FMRIPREP[2] nuisance regressors
- We fit voxel-wise L2-regularized (ridge) regression models using one of three datasets (ImageNet, COCO, Scenes).
- Features from an ImageNet-pretrained deep convolutional neural network (VGG-11) were used as predictors of voxel activity, subject to SVD prior to model fitting.
- An efficient leave-one-sample-out approach was used to select the optimal regularization strength on a per-voxel basis, using 80% of the given database to train/validate the model.
- R^2 and r^2 were used to assess model prediction (scale/shift-free and scaled/shifted)

BOLD5000

(Chang, Pyles, Marcus, Gupta, Tarr, and Aminoff, 2019)

Figure 3. Breakdown of stimuli in BOLD5000

Conclusions and Discussion

- Univariate mean response differences in high level vis. cortex supported an objects -> scenes continuum across BOLD5000 stimulus sets (Figs 3,4,5)
- Positive R^2 prediction in voxels with large univariate difference (Figs 1/4, 6) suggests that some of the univariate difference is due to graded variation along common representational dimensions/features
- Greater precedence of large r^2 prediction vs. R^2 prediction (Fig 6) in voxels with univariate differences may indicate a nonlinear scaling of similar features, e.g. a disproportionately large mean response to scenes in RSC/OPA/PPA
- Comparison of univariate and encoding model results suggests both overlap and divergence in representational mechanisms for objects and scenes in high level visual cortex
- Our method could also be applied as a special case of mixed RSA[5] in which the mixing is computed for multiple stimulus sets to compute RDMs for a common test set, which are correlated to test representational similarity

Encoding model results

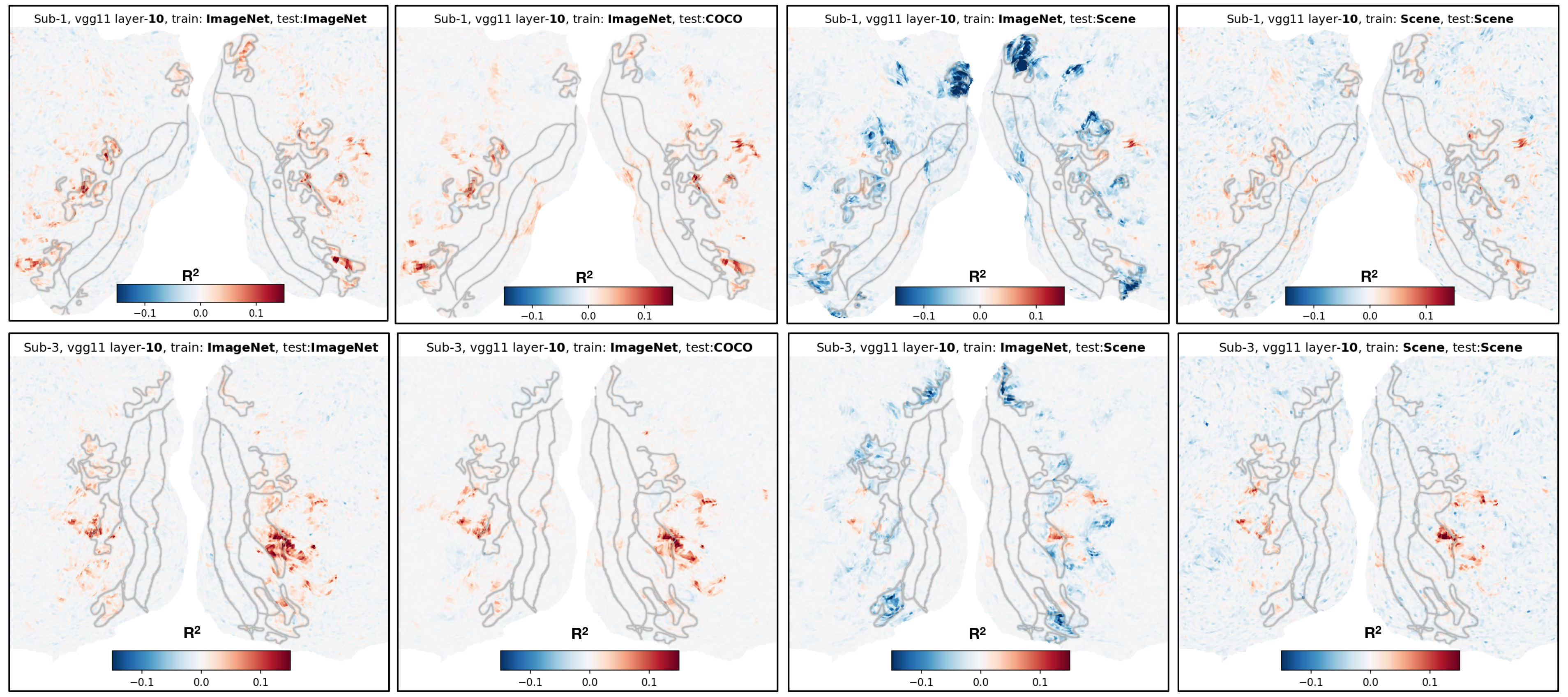

Figure 1. Whole-brain cross-database R^2 prediction for the penultimate layer of vgg11 in 2 example subjects

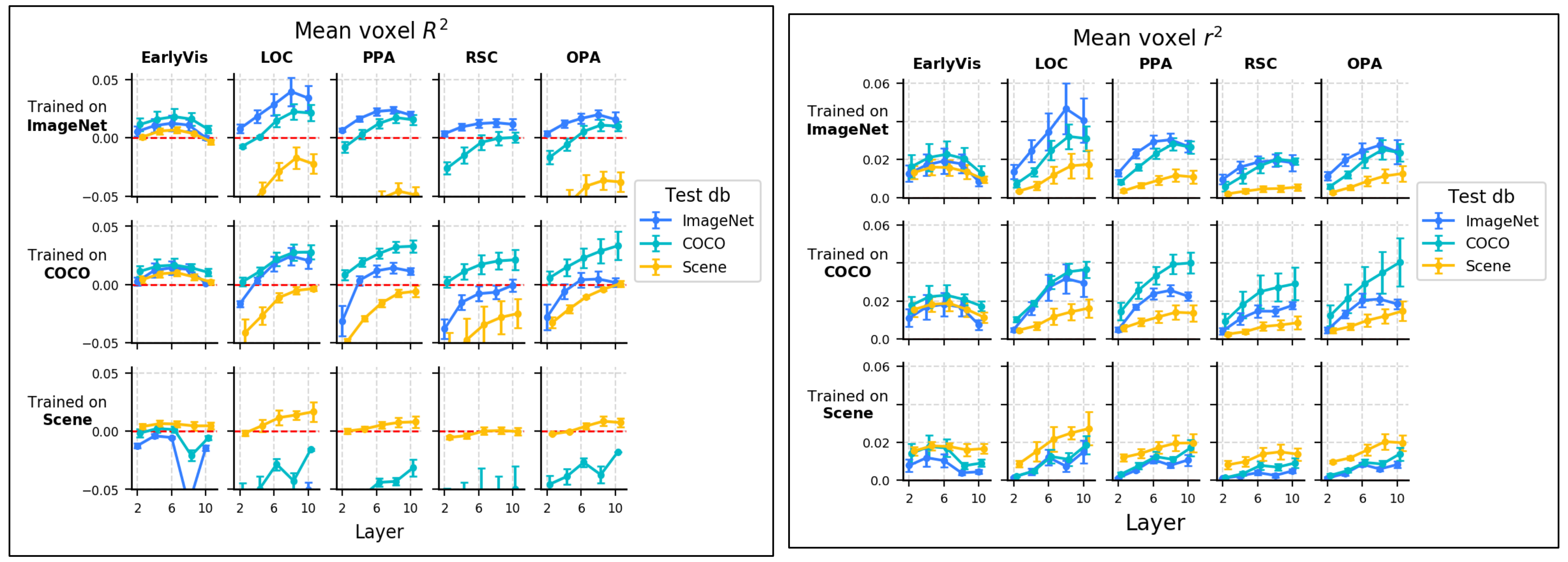

Figure 2. Encoding model generalization results in functionally-defined regions of interest, R^2 (left) and r^2 (right)

Univariate results

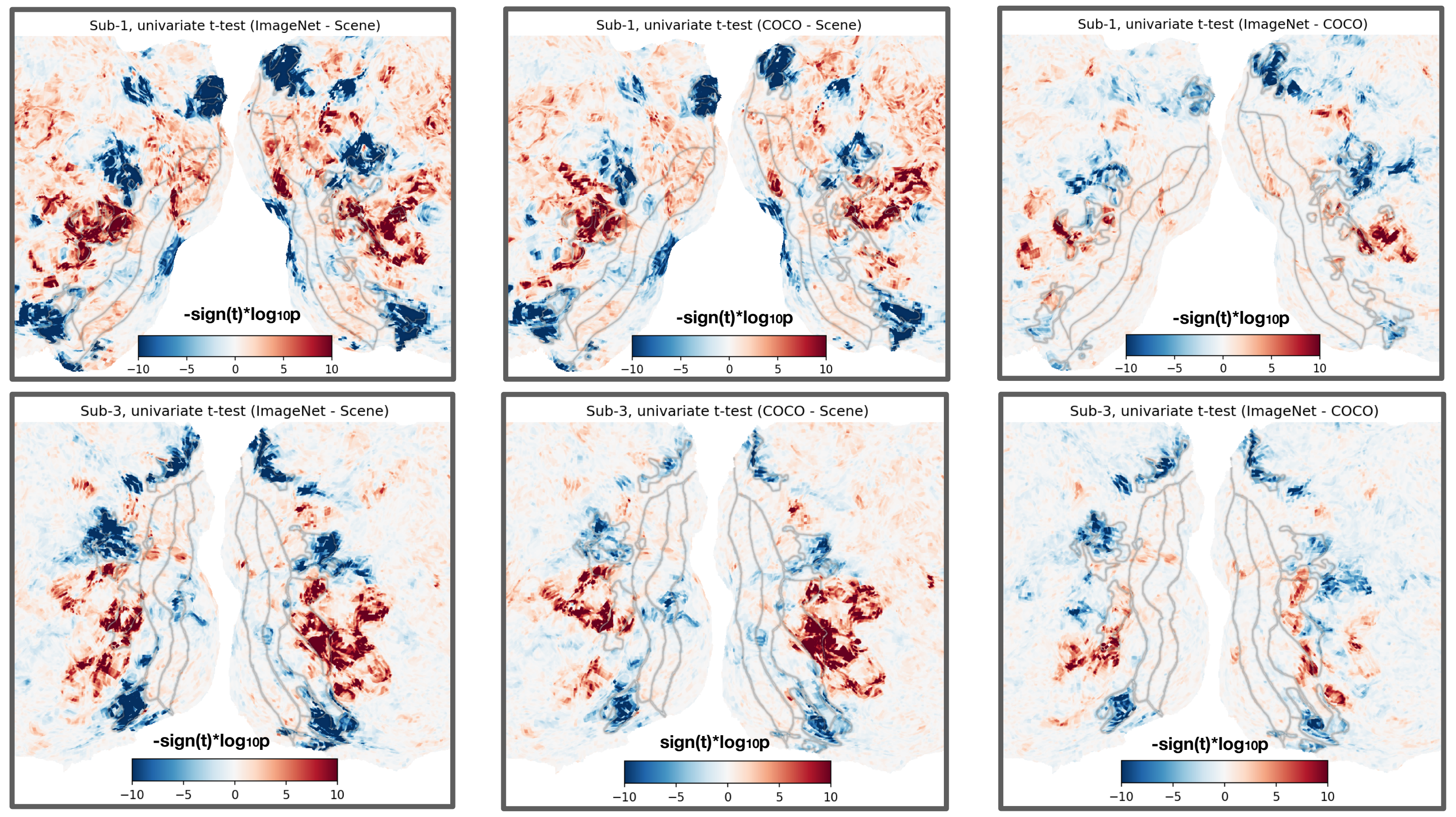

Figure 4. Univariate cross-database t-test significance maps for BOLD5000 in 2 example subjects

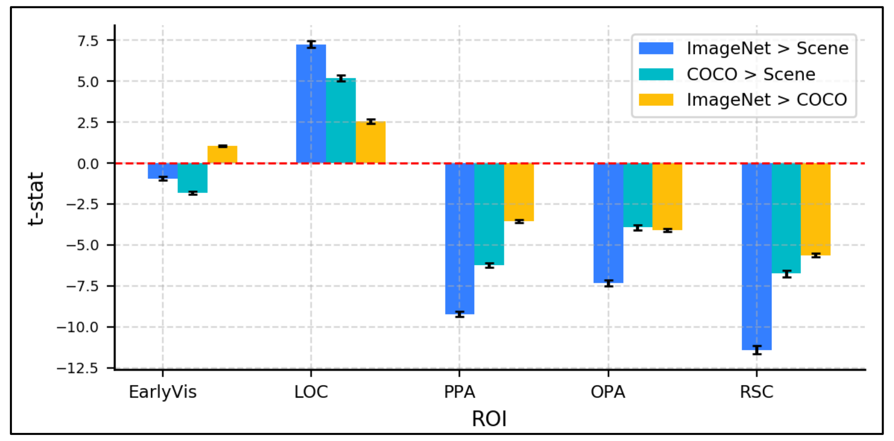

Figure 5. Mean voxel univariate t-test contrasts between datasets using 3 full subjects in BOLD5000

Acknowledgments

Special thanks to Nadine Chang, Elissa Aminoff, Mike Tarr and the authors of BOLD5000 for releasing a great data set to the public, and to the Tarr Lab for a helpful discussion about this work. We also thank the students of CMU Advanced Intro. to Machine Learning and instructor Nina Balcan for their feedback on an early portion of the project.

Comparing univariate and encoding model results

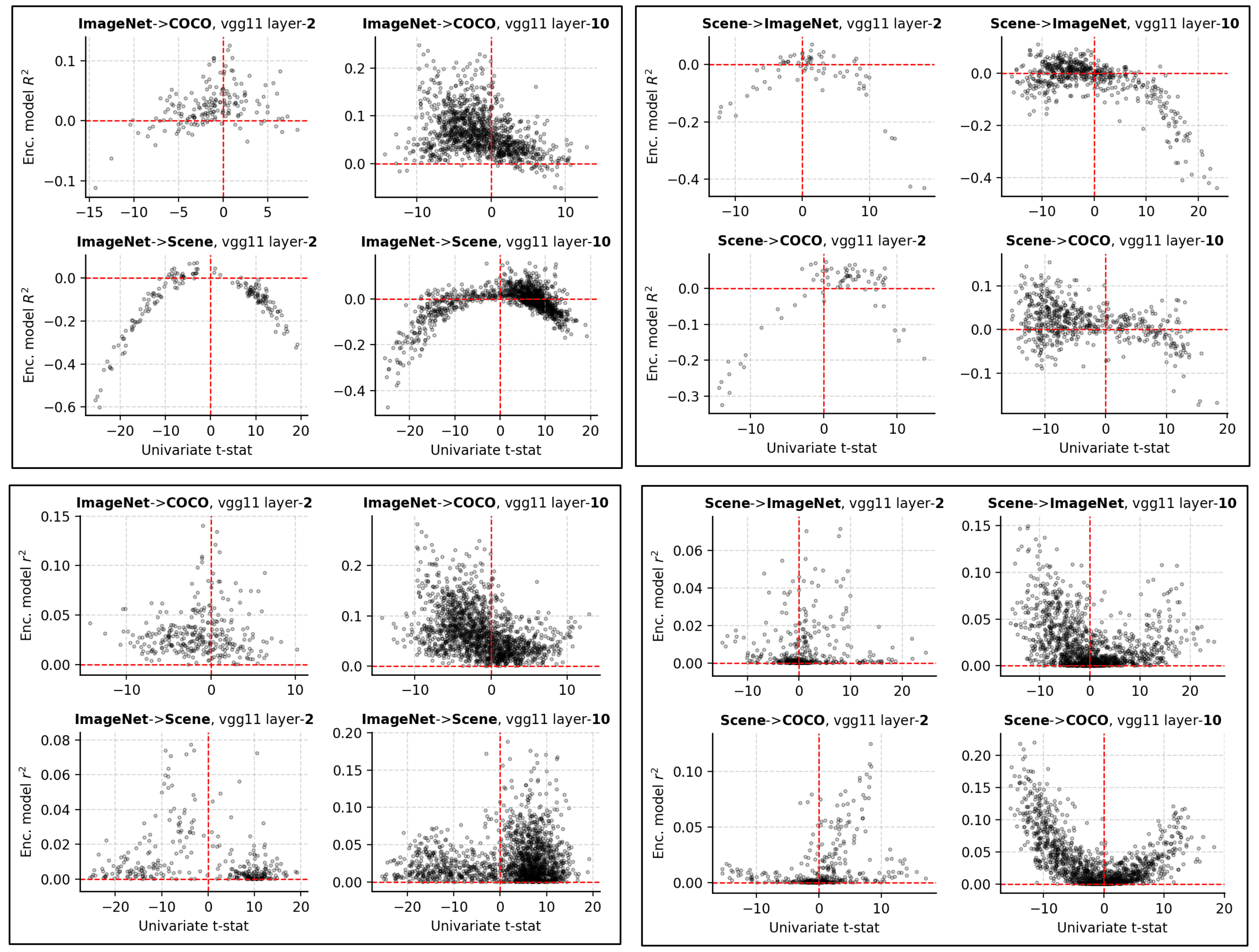

Figure 6. Subset of scatter plots of univariate t-stat vs. encoding model prediction across all voxels exceeding (R^2 or r^2> 0.05) for same-dataset generalization.

References

[1]Chang, N., Pyles, J. A., Marcus, A., Gupta, A., Tarr, M. J., & Aminoff, E. M. (2019). BOLD5000, a public fMRI dataset while viewing 5000 visual images. *Scientific Data*, *6*(49), 1–18.

[2]Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., … Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116.

[3]Gao, J. S., Huth, A. G., Lescroart, M. D., & Gallant, J. L. (2015). Pycortex: an interactive surface visualizer for fMRI. *Frontiers in Neuroinformatics*, *9*(September), 1–12.

[4]Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(November), 4.

[5]Khaligh-Razavi, S.-M., Henriksson, L., Kay, K., & Kriegeskorte, N. (2017). Fixed versus mixed RSA: Explaining visual representations by fixed and mixed feature sets from shallow and deep computational models. *Journal of Mathematical Psychology*, *76*, 184–197.