Do vision and imagery share common principal signal components?

Tiasha Saha Roy (tsaharoy@umn.edu)
Department of Neuroscience, University of Minnesota
Minneapolis, MN 55455 USA

Jesse Breedlove (jbreedlo@umn.edu)
Department of Psychology, University of Minnesota
Minneapolis, MN 55455 USA

Ghislain St-Yves (gstyves@umn.edu)
Department of Neuroscience, University of Minnesota
Minneapolis, MN 55455 USA

Kendrick Kay (kay@umn.edu)
Department of Radiology, University of Minnesota
Minneapolis, MN 55455 USA

Thomas Naselaris (nase0005@umn.edu)
Department of Neuroscience, University of Minnesota
Minneapolis, MN 55455 USA

Abstract

Most people experience mental imagery as an approximation to seeing. However, brain activity during acts of imagery typically exhibits lower signal-to-noise ratio (SNR) compared to vision, particularly in the early visual areas. In a previous work visual signal compression was suggested to be a plausible explanation for this apparent reduction of signal variance, and hence SNR, during imagery. In this work we explore the specific dimensions of visual coding that are preserved during imagery. We used an autoencoding voxel-to-voxel framework on data from a 7T fMRI imagery experiment to estimate the principal components of the visual and imagery signal spaces. The first principal imagery component separates imagery activity patterns in a similar way as the first principal visual component that separates visual activity patterns, implying that the same type of visual features are represented. However, visual variance is substantially lower along the imagery components, which is consistent with the compression hypothesis. Our results so far suggest that while the principal visual and imagery signal components exhibit apparent similarities in their coding for certain features, they do not seem to be completely identical.

Keywords: vision; mental imagery; fMRI; vox-to-vox models;

Introduction

Prior research (Naselaris, Olman, Stansbury, Ugurbil, & Gallant, 2015; Pearson, Naselaris, Holmes, & Kosslyn, 2015) indicates that the visual system might be using the same visual features to represent visual and mental images. However, the signal-to-noise ratio during vision far surpasses that during imagery, particularly in the early visual areas (Breedlove, St-Yves, Olman, & Naselaris, 2020). This along with the fact that mental images are represented with lower spatial resolution than seen images (Breedlove et al., 2020) led to the hypothesis that signal during imagery could be more precisely understood as a compressed form of the visual signal (Roy, Breedlove, St-Yves, Kay, & Naselaris, 2023). Using voxel-to-voxel predictive models (Mell, St-Yves, & Naselaris, 2021) in an fMRI mental imagery dataset it was found that a model that explicitly reduces dimensionality of visual activity on its way to predicting imagery activity fares better than one that doesn't (Roy et al., 2023). The compression model also revealed that the loss of dimensionality is more prominent in early visual areas, which is where a relative reduction in voxelwise SNR values was seen during imagery.

In this work we try to elucidate the relationship between the principal dimensions of the visual and imagery subspaces. We specifically ask 1) what proportion of the visual signal variance is preserved when visual signal is projected along the imagery dimensions and 2) whether the visual and imagery dimensions might encode similar visual features.

In order to answer these questions we use a vox-to-vox auto-encoding pipeline to estimate the principal dimensions of visual and imagery signal variance and project both visual and imagery data onto these specific dimensions.

Methods

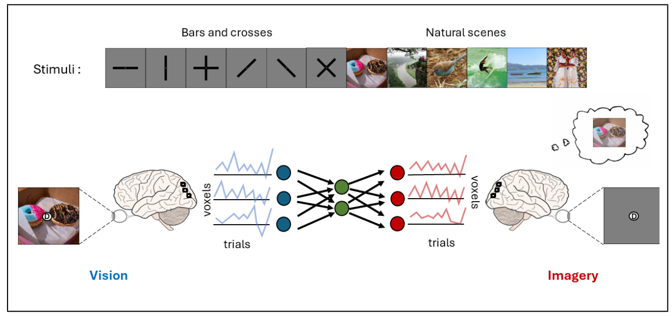

Figure 1: Top: Stimuli set used in NSD Imagery. Bottom: Schematic of voxel-to-voxel predictive model used to map between vision and imagery activity patterns from the NSD Imagery dataset.

Experiment: All 8 subjects of the Natural Scenes Dataset (NSD) experiment (Allen et al., 2022) took part in a separate scan session focusing on mental imagery. The stimuli set consisted of 4 bars, 2 crosses, 5 natural scenes, and 1 artwork (refer figure 1 Top). Participants viewed and in separate runs imagined these stimuli. Each stimulus was repeated 8 times, resulting in a total of 96 trials for each run type (vision/imagery).

ROI selection: We considered the visual ROIs V1, V2, V3, hV4—defined on the basis of an independent retinotopic mapping experiment—and three higher-level ROIs named 'ventral', 'lateral', and 'parietal'—defined on the basis of anatomical location. Voxels were selected based on their signal-to-noise ratio during the NSD experiment.

Voxel-to-voxel predictive models: We trained voxel-to-voxel linear regression models (Mell et al., 2021) where the activity in all voxels during vision predicts activity in those voxels during imagery (vis2img model). Models are trained individually for each NSD subject using a 4-fold cross-validation procedure. It is assumed that the transformation from visual representations to imagery representations takes place via a lower dimensional bottleneck (denoted by green dots in the illustration in figure 1 Bottom). The number of nodes in the bottleneck is the optimal 'rank' in this case and is estimated from the validation set using grid search over a range of possible ranks. We also train a separate vision-to-vision transformation (vis2vis model) where one trial of a particular seen image predicts another trial of the same image. The vision dimensions are estimated using this vis2vis model while the imagery dimensions are estimated from the vis2img model.

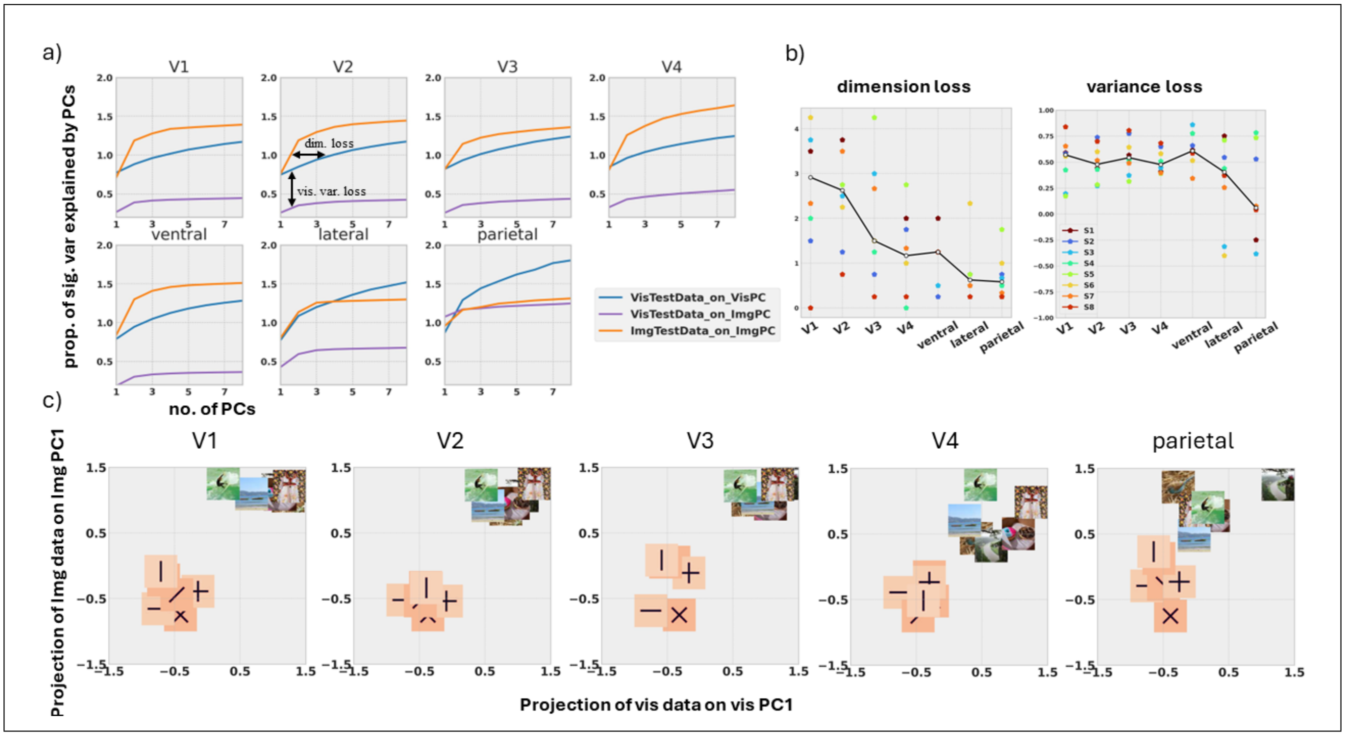

Figure 2: a) Cumulative signal variance explained by vision and imagery PCs. Variance explained is computed relative to the total variance explained when test data (vision/imagery) is projected onto the signal dimensions only. At the PC corresponding to the estimated signal dimension the cumulative variance explained reaches 1. Beyond that, all dimensions explain noise variance according to our model. Data has been averaged across all cross-validation folds and all NSD subjects. b) Left: Median loss in dimensionality from vision to imagery. Right: Median visual variance lost due to projection of visual data onto the imagery subspace in all ROIs. Medians are computed across 8 NSD subjects. c) Projection of trial-averaged brain activity on principal components of visual and imagery subspaces. Visual data has been projected onto the first visual PC and imagery data has been projected onto the first imagery PC.

Results

As reported in a prior work (Roy et al., 2023), the imagery subspace has fewer dimensions than the visual subspace, specifically for the early visual areas (refer to orange and blue curves in figure 2a). When visual activity from the test set is projected onto the imagery signal dimensions (estimated from the training and validation set), there seems to be a substantial loss of visual variance (denoted by the vertical distance between the blue and purple curve along the imagery signal dimension in figure 2a), hinting at a possible difference between the visual and imagery PCs. This difference seems consistently high across different visual areas where the difference in dimensionality was high (figure 2b). On projecting the trial averaged data from the 12 stimuli from vision and imagery onto their respective signal dimensions, we observe that the first visual PC sorts the visual data in a similar way as the first imagery PC in almost all ROIs (figure 2c), implying coding for similar visual features. In our current dataset the projections tend to cluster depending on the nature of stimuli used - synthetic vs naturalistic.

Discussion

We used a vox-to-vox modeling approach to extract the principal dimensions of visual and imagery signal spaces. A substantial reduction in visual signal variance was observed when vision trials are projected onto the Imagery PCs. This finding provides further evidence in support of the compressed vision hypothesis proposed in earlier work (Roy et al., 2023). However, the primary dimension of variance in both visual and imagery spaces sorts the images of the NSD Imagery dataset with respect to similar visual features—synthetic vs naturalistic. These results provide some evidence that even though the principal dimensions of visual and imagery signal code for similar visual features, they are not closely aligned in the space of brain activity. It is possible that this “misalignment” of seen and mental imagery dimensions minimizes the impact of mental imagery on seeing.

Acknowledgments

This work was supported by NIH R01EY023384 (T.N.). Collection of the NSD dataset was supported by NSF IIS-1822683 and NSF IIS-1822929.

References

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., ... others (2022). A massive 7t fmri dataset to bridge cognitive neuroscience and artificial intelligence. *Nature neuroscience*, *25*(1), 116–126.

Breedlove, J. L., St-Yves, G., Olman, C. A., & Naselaris, T. (2020). Generative feedback explains distinct brain activity codes for seen and mental images. *Current Biology*, *30*(12), 2211–2224.

Mell, M. M., St-Yves, G., & Naselaris, T. (2021). Voxel-to-voxel predictive models reveal unexpected structure in unexplained variance. *NeuroImage*, *238*, 118266.

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *Neuroimage*, *105*, 215–228.

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: functional mechanisms and clinical applications. *Trends in cognitive sciences*, *19*(10), 590–602.

Roy, T. S., Breedlove, J., St-Yves, G., Kay, K., & Naselaris, T. (2023). Mental imagery: Weak vision or compressed vision? Conference on Cognitive Computational Neuroscience. doi: 10.32470/CCN.2023.1693-0