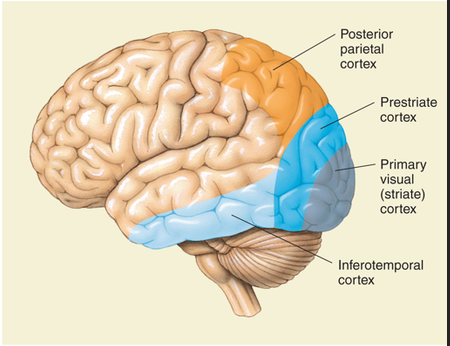
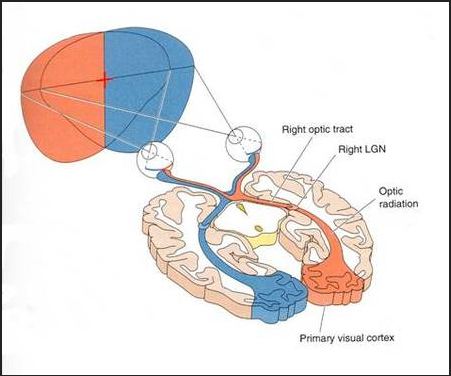
Performance-optimized hierarchical models predict neural responses in higher visual cortex

**Visual Cortex:**



The visual cortex of the [brain](https://en.wikipedia.org/wiki/Brain) is the area (area V1) of the [cerebral cortex](https://en.wikipedia.org/wiki/Cerebral_cortex) that processes [visual information](https://en.wikipedia.org/wiki/Visual_perception). The primary purpose of the visual cortex is to receive, segment, and integrate visual information. The processed information from the visual cortex is subsequently sent to other regions of the brain to be analyzed and utilized.

It is located in the [occipital lobe](https://en.wikipedia.org/wiki/Occipital_lobe). Sensory input originating from the [eyes](https://en.wikipedia.org/wiki/Eye) travels through the [lateral geniculate nucleus](https://en.wikipedia.org/wiki/Lateral_geniculate_nucleus) in the [thalamus](https://en.wikipedia.org/wiki/Thalamus) and then reaches the visual cortex.

Both [hemispheres of the brain](https://en.wikipedia.org/wiki/Cerebral_hemisphere) include a visual cortex; the visual cortex in the left hemisphere receives signals from the right [visual field](https://en.wikipedia.org/wiki/Visual_field), and the visual cortex in the right hemisphere receives signals from the left visual field. Representation and recognition of objects are thought to be functions of higher extrastriate cortical areas. The primary visual cortex is the most studied visual area in the brain. In mammals, it is located in the posterior pole of the occipital lobe and is the simplest, earliest cortical visual area. It is highly specialized for processing information about static and moving objects and is excellent in pattern recognition.

The visual cortex subdivides into five different areas based on structural and functional classifications. The hypothesis is that as visual information gets passed along, each subsequent cortical area is more specialized than the last. Neurons in the visual cortex often respond to stimuli within a fixed receptive field, the area of the visual field that they respond to, and the neurons in each visual area respond to different types of stimuli. One of the best-studied examples of specialized cells is that of simple and complex cells. Simple cells, which are found mostly in V1, respond to specific types of visual cues such as the orientation of edges and lines. Complex cells, which occur in V1-V3, are like simple cells in that they respond to edges and orientations, but they do not appear to represent a single receptive field. Instead, they respond to the summation of several receptive fields that become integrated from many simple cells. Also, complex cells respond preferentially to movement in specific directions. Other examples of specialized cells include end-stopped cells, which detect line endings, and bar and grating cells.[[3]](https://www.ncbi.nlm.nih.gov/books/NBK482504/#)



**Thalamus:**

The thalamus is a mostly gray matter structure of the diencephalon that has many essential roles in human physiology. The thalamus is composed of different nuclei that each serve a unique role, ranging from relaying sensory and motor signals, as well as regulation of consciousness and alertness.While the thalamus is mostly gray matter (cell bodies of neurons), there are some areas of white matter (axons). The external and internal medullary laminae are white matter structures of the thalamus. The external medullary laminae cover the lateral surface of the thalamus, and the internal medullary laminae divide the thalamic nuclei into anterior, medial, and lateral groups

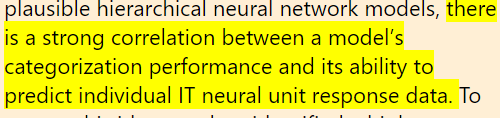
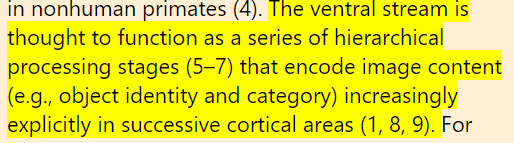
**Inferior temporal (IT) cortex:**

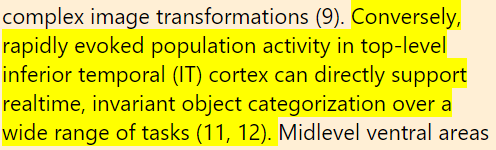
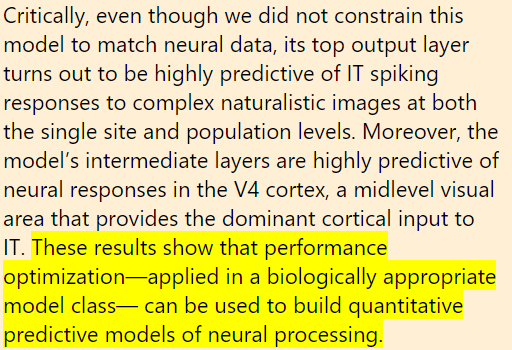
The inferior temporal (IT) cortex plays a critically important role in the visual recognition of objects. The visual recognition system in the IT cortex is distributed in multiple areas, including area TE and the rhinal cortex.

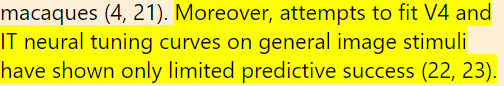
Lesions in the IT cortex cause behavioural changes that depend on the reward schedule. In one study, monkeys learned a task in which visual cues indicated the number of successful trials (one, two, or three) required to receive a reward. Monkeys performed each trial differently depending on the reward schedule, but lesioning or suppressing the expression of [dopamine D2 receptors](https://www.sciencedirect.com/topics/neuroscience/dopamine-receptor-d2) in the perirhinal cortex abolished the reward-schedule-dependent behaviour.

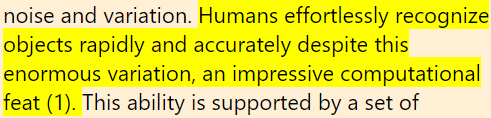
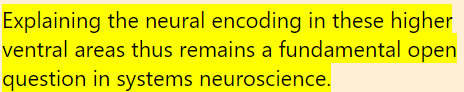
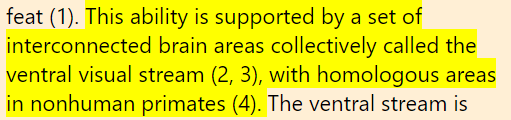
Summary Paper:

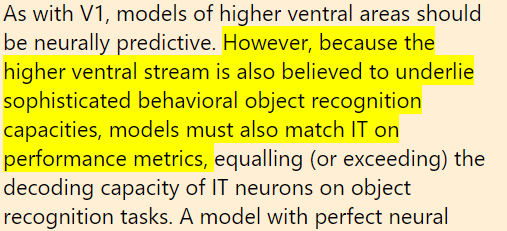
Keynotes:

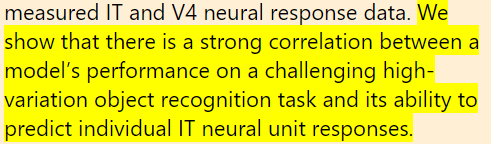
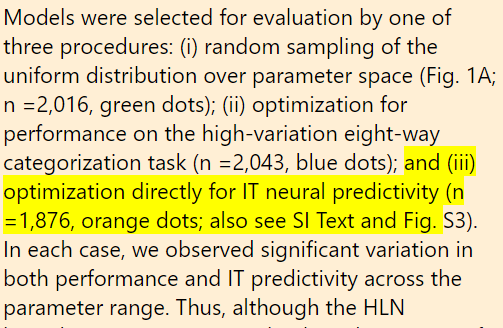


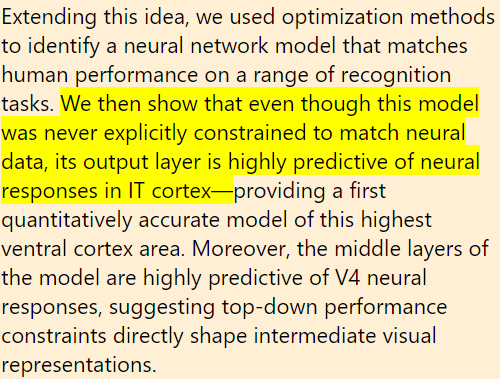


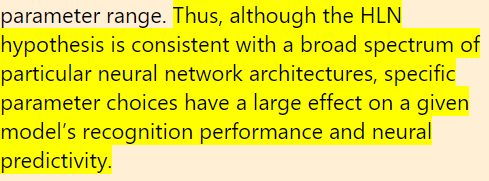


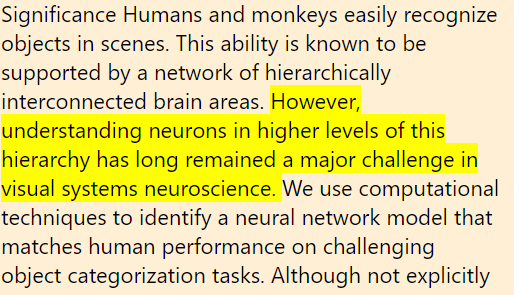
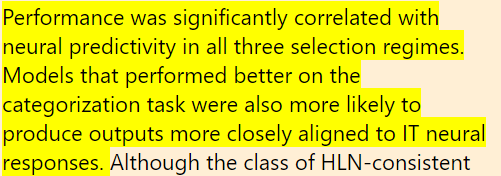


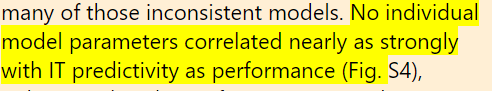


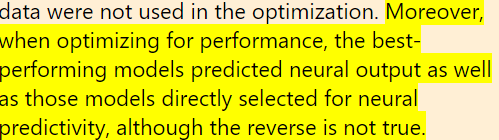


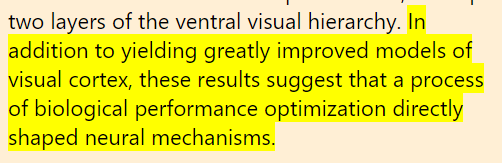


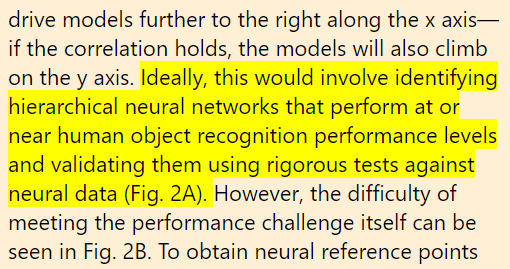


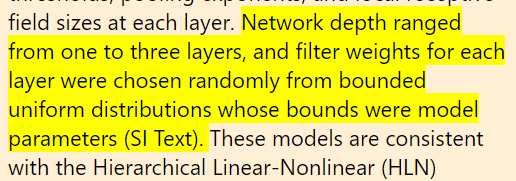


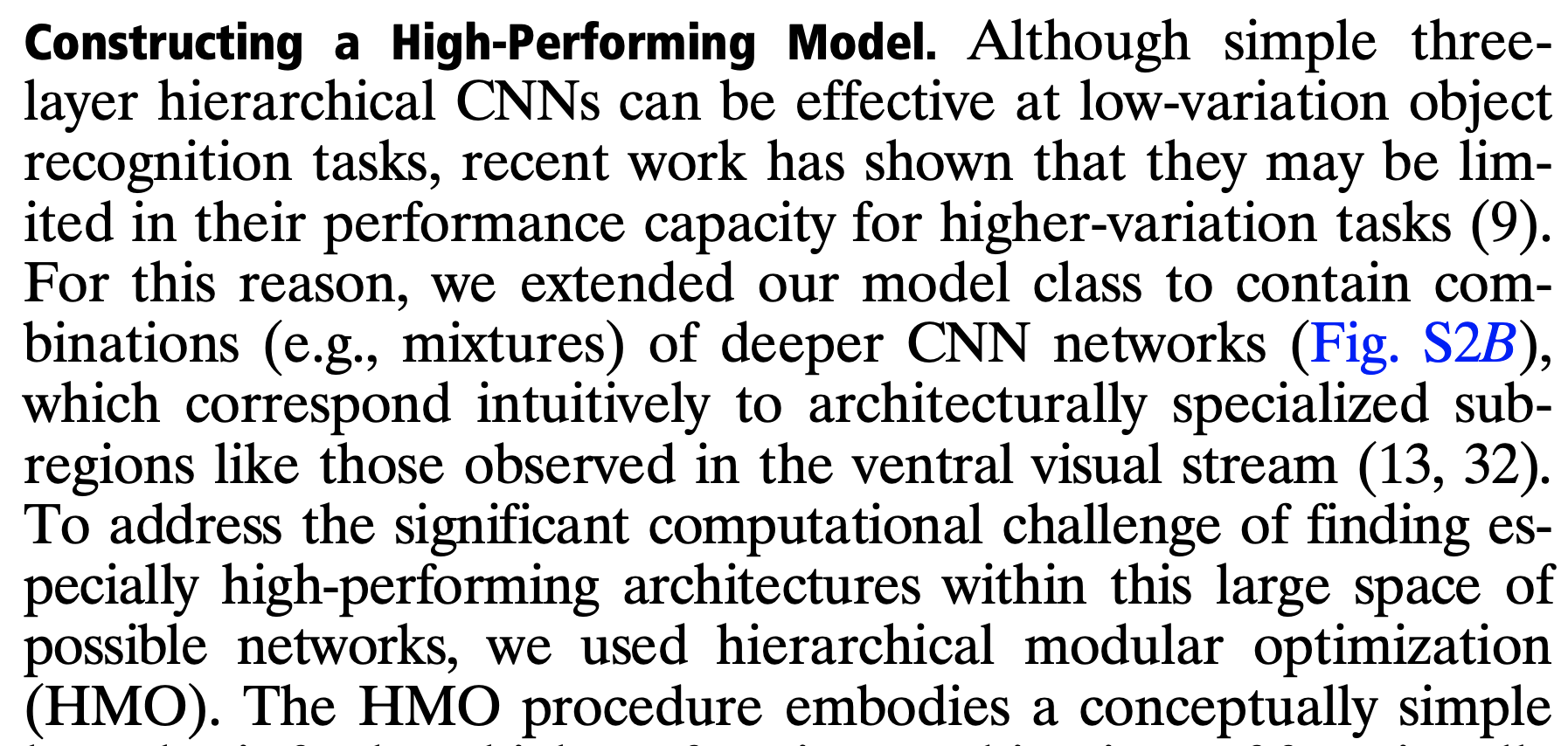
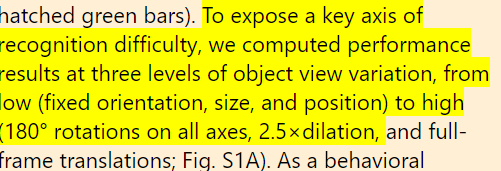


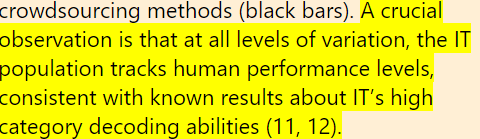


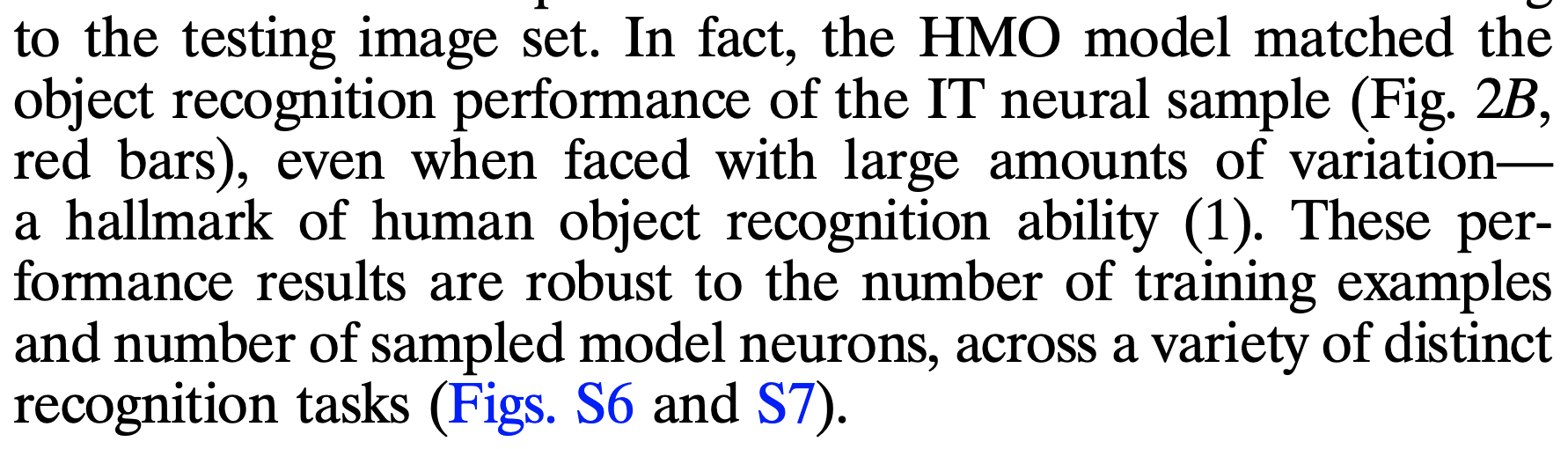


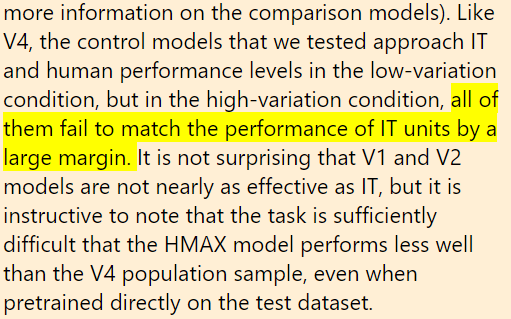


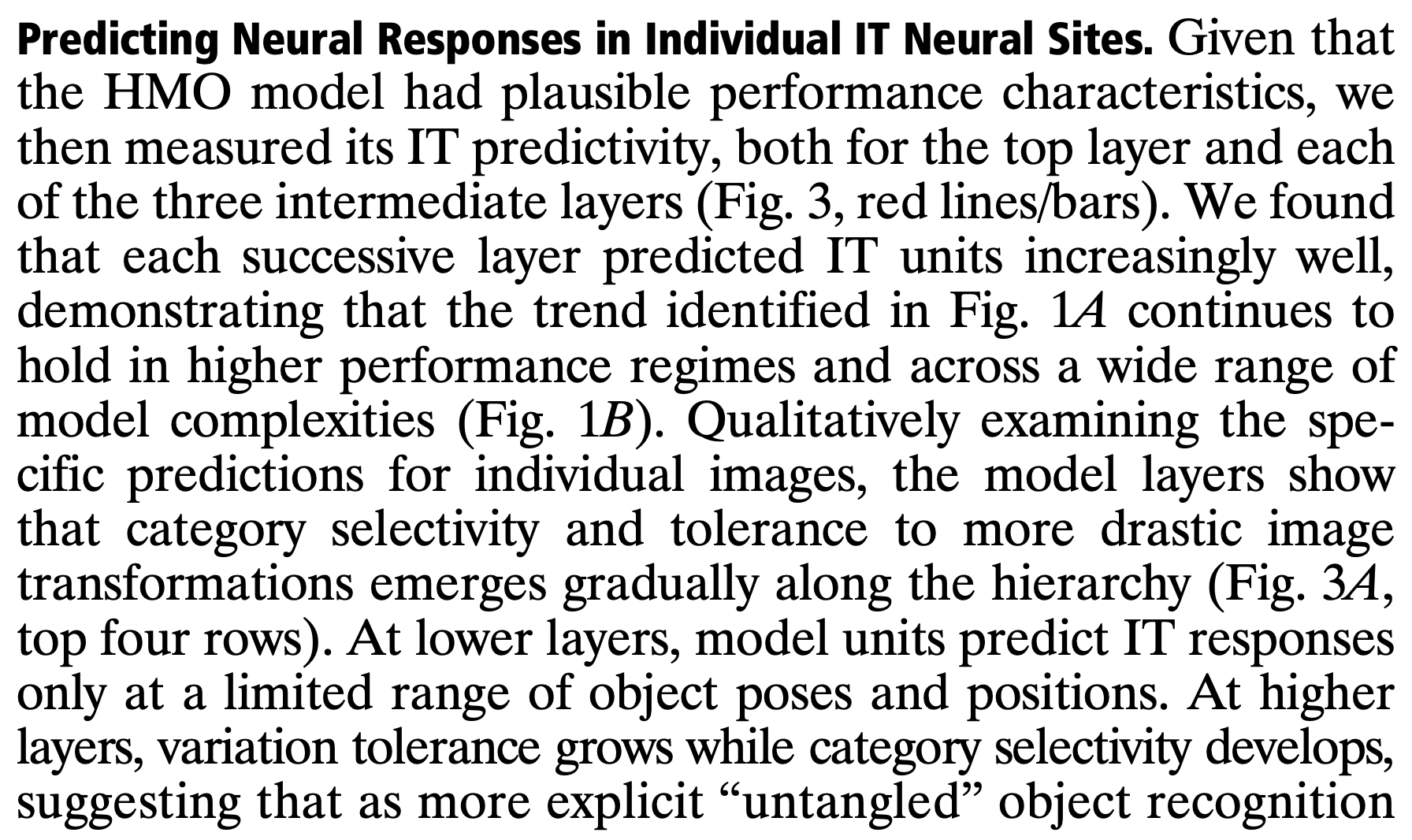


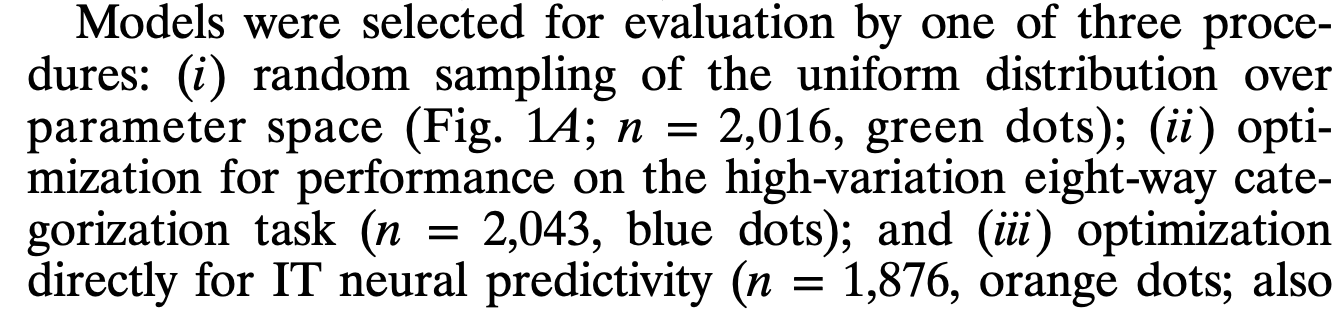
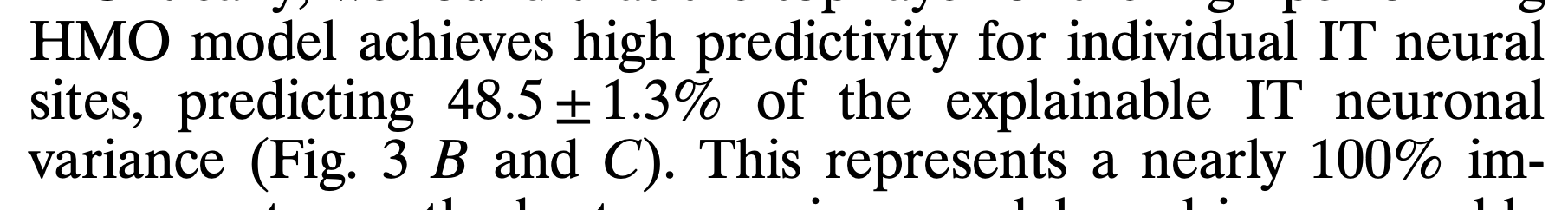
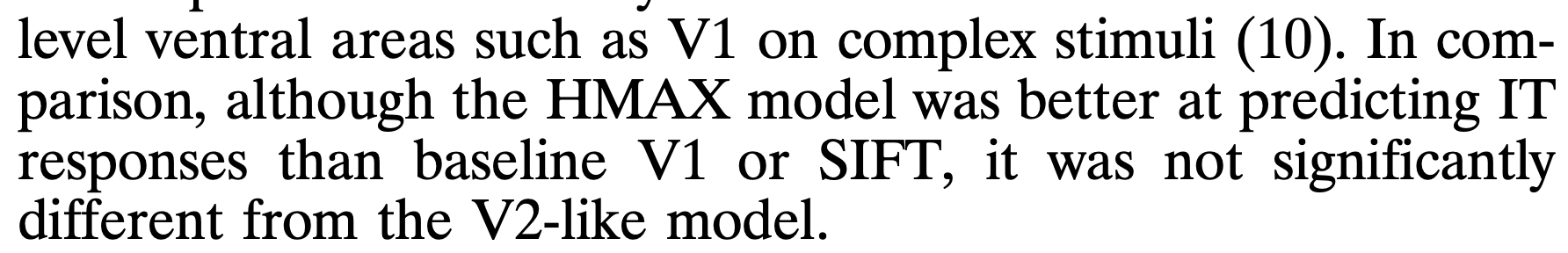


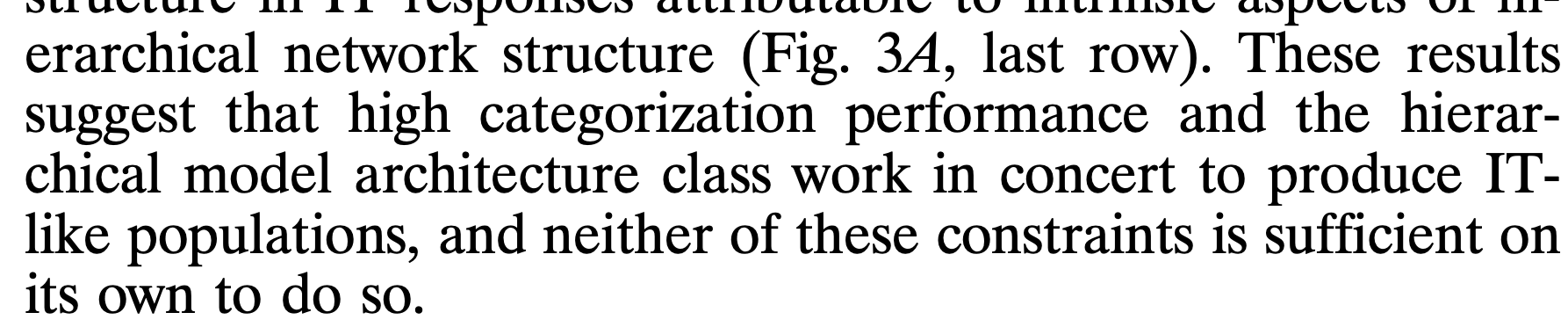
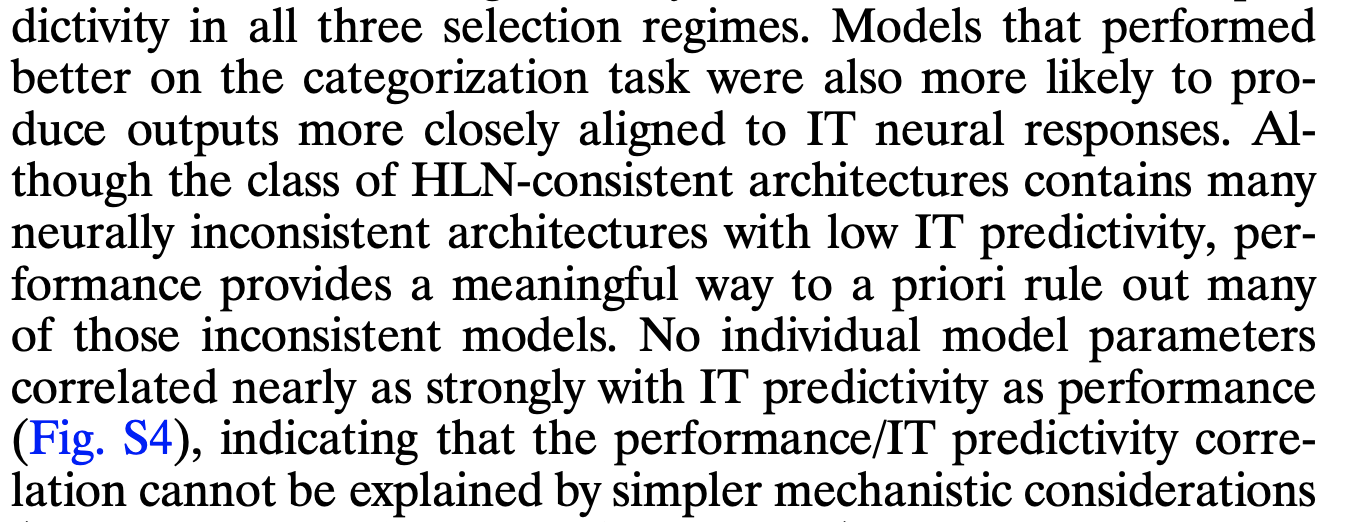


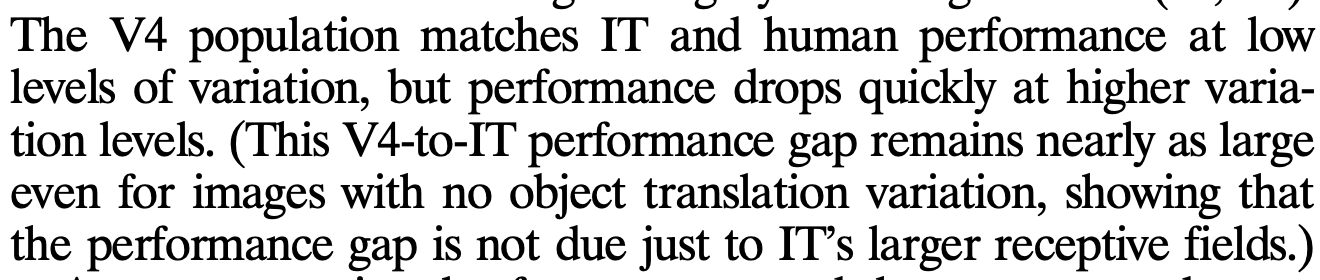


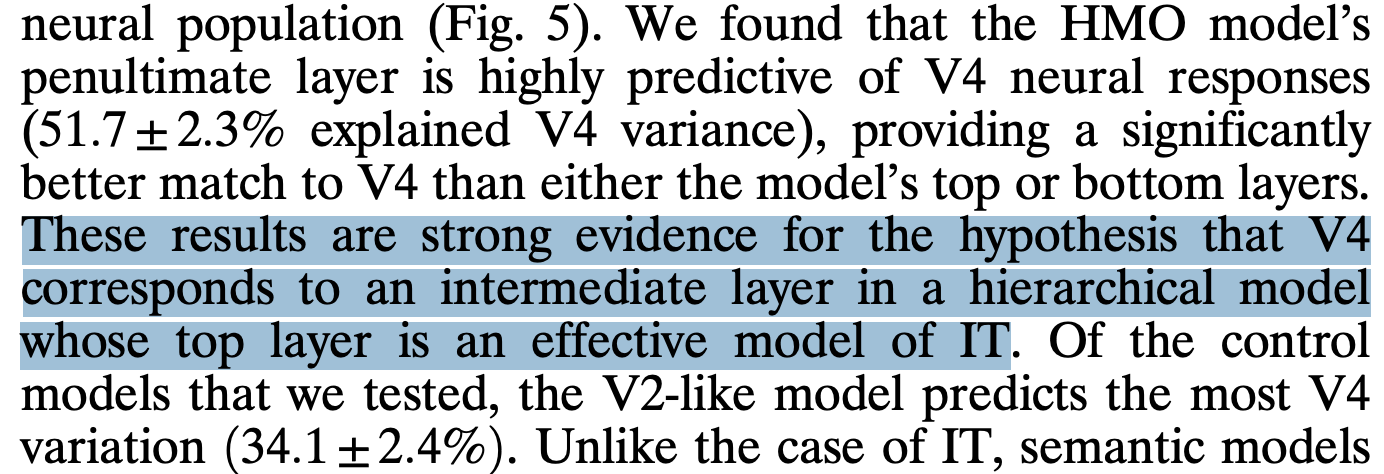












Summery external:

Humans and monkeys easily recognize objects in scenes. This ability is known to be supported by a network of hierarchically interconnected brain areas. However, understanding neurons in higher levels of this hierarchy has long remained a major challenge in visual systems neuroscience. We use computational techniques to identify a neural network model that matches human performance on challenging object categorization tasks. Although not explicitly constrained to match neural data, this model turns out to be highly predictive of neural responses in both the V4 and inferior temporal cortex, the top two layers of the ventral visual hierarchy. In addition to yielding greatly improved models of visual cortex, these results suggest that a process of biological performance optimization directly shaped neural mechanisms.

**Summery Idan**

Significant research and effort has been made in understanding the lower Ventral areas such as V1, however, higher ventral cortical areas, such as V4, V5 and IT are much more difficult for humans to comprehend. Thus explaining those encoding in the higher ventral areas remind an open question for neuroscience and the world in general.

We found a way to tackle this:

Instead of following the path and the input manipulation through the ventral processing in lower processing areas, in order to understand the higher areas. Their approach is by measuring the neuron behaviour and drawing conclusions from that. More precisely, they use computational techniques to identify a neural network model that matches human performance on object categorization and assess them against measured IT and V4 neural response data. This method provides the first quantitatively accurate model of higher ventral cortical area as well as accurately V4 neural responses.

**But How?**

**Base case ⇒ Humans:**

They created a test case where they measured the IT and V4 responses on 5,760 images of photorealistic 3D objects drawn from different categories such as animals, boats, cars, chairs, faces etc. To make the recognition harder for the task, they played with the variation (scale,position ,etc). As we know that these processes are relatively easy for humans to perform but harder for machines. By placing the object on a randomly selected cluttered natural scene they ensured uncorrelation between the object identity and background.They collected 168 IT neurons for each image.

**Test:**

They compared this classification task on various models as on 3 main categories of Variation task:

* Low variation : for instance, where all the faces are represented from the front
* middle variation : where the complexity of the object representation has increased ( more difficult angle etc…)
* High variation task: the highest complexity level in this research.

It is important to note that many object recognition models have a higher low variation prediction.

The models which were checked are SVM with linear regression classifier, Hmax, V2 like, V1 like Sift, Pixel , HMO, humans, and IT population and V4 population.

They obviously found a strong correlation between how well the human could classified correctly, to its neuron IT measured firing.

An interesting result is that they found that in a high variation task, HMO, which stands for hierarchical modular optimization ( a more optimized version of CNN) had very high similarity to the actual IT population structure, close to the split-half noise ceiling of the IT population.These results are strong evidence for the hypothesis that V4 corresponds to an intermediate layer in a hierarchical model whose top layer is an effective model of IT.

Together these results suggest that performance optimization not only drives top-level output model layers to resemble IT, but also imposes biologically consistent constraints on the intermediate feature representations that can support downstream performance.

They found that the neural population predicted by the output layer of the HMO model had very high similarity to the actual IT population structure, close to the split-half noise ceiling of the IT population

At first, they built a base SVM model and developed it with a linear classifier and performed cross validation. The models were drawn from a large dimension space of CNN imitating the brain pipeline processes.

HLN( hierarchical linear nonlinear hypothesis that higher level neurons output a linear weighing of input from intermediate-level. Which means that neurons followed by a simple additional non linearity. Although performance was significantly correlated with neural predictivity in all three selection regimes, HLN-consistent architectures contain many neurally inconsistent architectures with low IT predictivity. But it did give a good indication of the future.

They built various models to tackle this approach such as SVM with linear regression classifier, Hmax, V2 like, V1 like Sift, Pixel , HMO. by improving and optimizing in each step of the model, they measure prediction on different kinds of input. While low variation models are much less complicated to perform, as we increase the medium

intermediate-level

After evaluating

. Starting with SVM model

Result :

They use representation dissimilarity matrix (RDM) which measures the IT population, when images are ordered by category.

We found that the neural population predicted by the output layer of the HMO model had very high similarity to the actual IT population structure, close to the split-half noise ceiling of the IT population (Fig. 4B). This implies that much of the residual variance unexplained at the single-site level may not be relevant for object recognition in the IT population level code.

These results are strong evidence for the hypothesis that V4 corresponds to an intermediate layer in a hierarchical model whose top layer is an effective model of IT.

Unlike the case of IT, semantic models explain effectively no variance in V4, consistent with V4’s lack of category selectivity. Together these results suggest that performance optimization not only drives top-level output model layers to resemble IT, but also imposes biologically consistent constraints on the intermediate feature representations that can support downstream performance.

Performance was significantly correlated with neural predictivity in all three selection regimes. Models that performed better on the categorization task were also more likely to produce outputs more closely aligned to IT neural responses.

Check:

It is not surprising that V1 and V2 models are not nearly as effective as IT, but it is instructive to note that the task is sufficiently difficult that the HMAX model performs less well than the V4 population sample, even when pre trained directly on the test dataset.

Constructing a High-Performing Model. Although simple three layer hierarchical CNNs can be effective at low-variation object recognition tasks, recent work has shown that they may be limited in their performance capacity for higher-variation tasks (9). For this reason, we extended our model class to contain combinations (e.g., mixtures) of deeper CNN networks (Fig. S2B), which correspond intuitively to architecturally specialized subregions like those observed in the ventral visual stream (13, 32). To address the significant computational challenge of finding especially high-performing architectures within this large space of possible networks, we used hierarchical modular optimization (HMO).

Algorithmically, HMO is analogous to an adaptive boosting procedure (33) interleaved with hyperparameter optimization (see SI Text and Fig. S2C). As a pre training step, we applied the HMO selection procedure on a screening task (Fig. S1B). Like the testing set, the screening set contained images of objects placed on randomly selected backgrounds, but used entirely different objects in totally nonoverlapping semantic categories, with none of the same backgrounds and widely divergent lighting conditions and noise levels.

Using the same classifier training protocol as with the neural data and control models, we then tested the HMO model to determine whether its performance transferred from the screening to the testing image set. In fact, the HMO model matched the object recognition performance of the IT neural sample (Fig. 2B, red bars), even when faced with large amounts of variation— a hallmark of human object recognition ability (1). These performance results are robust to the number of training examples and number of sampled model neurons, across a variety of distinct recognition tasks The representation dissimilarity matrix (RDM) is a convenient tool comparing two representations on a common stimulus set in a task-independent manner (4, 35). Each entry in the RDM corresponds to one stimulus pair. When images are ordered by category, the RDM for the measured IT neural population (Fig. 4A) exhibits clear block-diagonal structure.We found that the neural population predicted by the output layer of the HMO model had very high similarity to the actual IT population structure, close to the split-half noise ceiling of the IT population (Fig. 4B). This implies that much of the residual variance unexplained at the single-site level may not be relevant for object recognition in the IT population level code.

These results are strong evidence for the hypothesis that V4 corresponds to an intermediate layer in a hierarchical model whose top layer is an effective model of IT.

Unlike the case of IT, semantic models explain effectively no variance in V4, consistent with V4’s lack of category selectivity. Together these results suggest that performance optimization not only drives top-level output model layers to resemble IT, but also imposes biologically consistent constraints on the intermediate feature representations that can support downstream performance.

Tests:

* We first measured IT neural responses on a benchmark testing image set that exposes key performance characteristics of visual representations