Title

Evidence that feedback is required for object identity inferences computed by the ventral stream

Author List

Kohitij Kar*, Jonas Kubilius, Elias Issa, Kailyn Schmidt and James J. DiCarlo

Summary

The primate ventral visual stream for object recognition contains prominent cortico-cortical feedback connections. However, the most accurate models of online, rapid (<200 ms) inference in the ventral stream are largely feedforward (hierarchical convolutional neural networks, HCNN). Might the appropriate inclusion of feedback connections in those models improve their explanatory power? We reasoned that, the impact of feedback connections would be most easily revealed in neural population activity at the top of the ventral visual hierarchy (inferior temporal cortex, IT), because the IT representation benefits from feedback connections along the entire hierarchy.

Because prior work shows that linear decoders accurately model IT's estimate of object identity, we could look for a neural signature that would imply a computationally-critical role of feedback in online inference. Specifically, we hypothesized that, for images that require feedback circuits to resolve objects, IT's estimate of object identity should emerge later (relative to other images). To discover such images, we behaviorally tested both synthetic images and photographs to obtain two groups of images — those for which object identity is easily extracted by the primate brain, but not solved by HCNNs ("challenge" images), and those that primates and models easily solve (control images). We then recorded IT population activity in two monkeys while they performed core object identity estimation (100 ms viewing) on each image (1360 images, 10 possible objects, randomly interleaved to neutralize attention).

We found that, in both monkeys, IT's solution (linear decode) for challenge images took ~20 msec longer to emerge than control images. This difference could not be explained by differences in neural latency, firing rates, or low-level image properties such as contrast. These results imply the importance of feedback in ventral stream object inference, and the image-by-image differences constrain the next generation of ventral stream models.

Additional Details

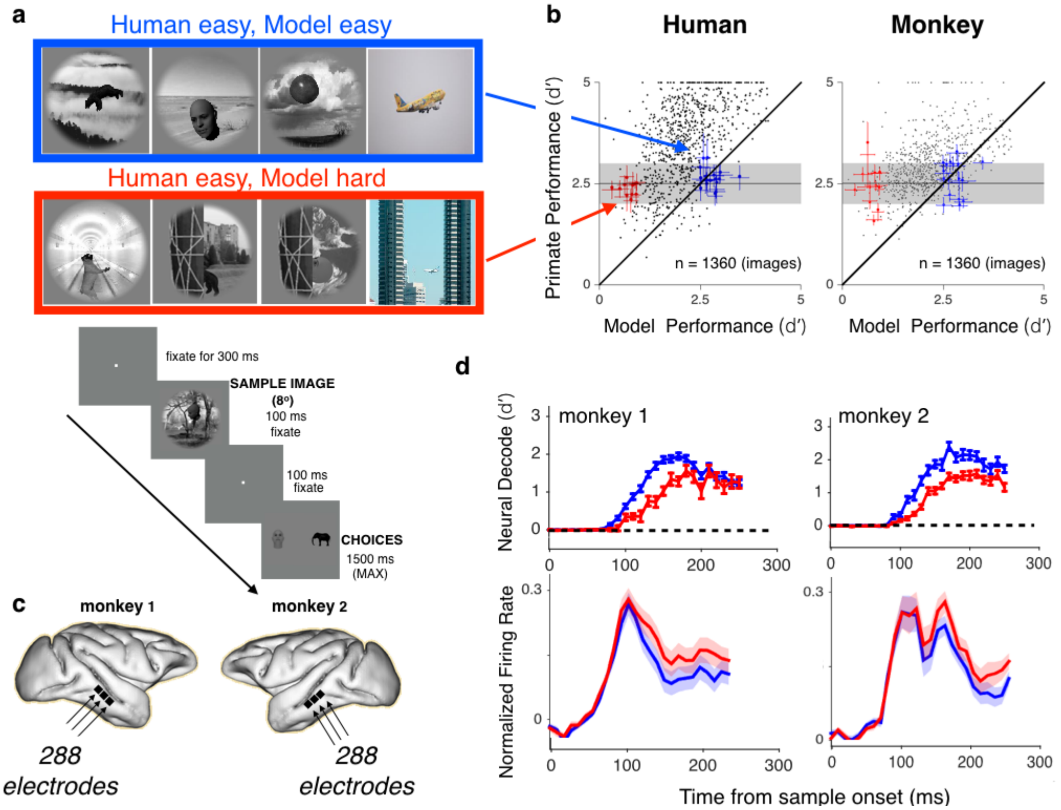

Figure1. Experimental design and results. **a**) Top panel: examples of “challenge” (red) and control (blue) images (of 1360 total tested). Bottom: timeline of events on each trial. Subjects fixate a dot. A sample image (8 deg) containing one of ten possible objects was shown for 100 ms. After a 100 ms delay, a target object and a distractor object (from the other 9 objects). appeared, and the monkey indicated which object was in the sample image by making a saccade to one of the two choices. **b**) Left panel: comparison of human performance (pooled data from 86 amazon mechanical turk workers) with performance of a HCNN model (Alex Net). Each dot shows performance on one image (averaged over all 9 distractor tasks). Right panel: comparison of monkey performance (trials pooled across two monkeys) to that same HCNN model. We compared images that had the same human difficulty (gray band), but very different model difficulty (red vs. blue dots). The red and blue dots representing the same images are plotted on the right panel demonstrating similarity between human and monkey behavior. Note- for many images primates and model perform similarly (dots along black unity line). However, primates outperformed the model for many other images (dots on the top left quadrant of the figure). **c)** Two macaques were each implanted with 3 Utah arrays (each with 96 electrodes) in IT. **d**) Top panels: neural decoder (linear kernel SVM) performance on held out sample images using the IT population as a feature basis (Monkey 1, 192 IT multi-unit sites; Monkey 2, 191 multi-unit sites), as a function of time from sample image onset. Red and blue colors show the mean performance over all challenge (red) and control (blue) images. Neural solutions for the challenge images emerge ~20 ms later than the control images. Averaged population firing rates for both monkeys (bottom panels) show no differences in latency between the challenge (red) and control (blue) images. Error bars and shading are SEM.