

## Supplementary Information for

### Face neurons encode non-semantic features

Alexandra Bardon, Will Xiao, Carlos R. Ponce, Margaret S. Livingstone, and Gabriel Kreiman

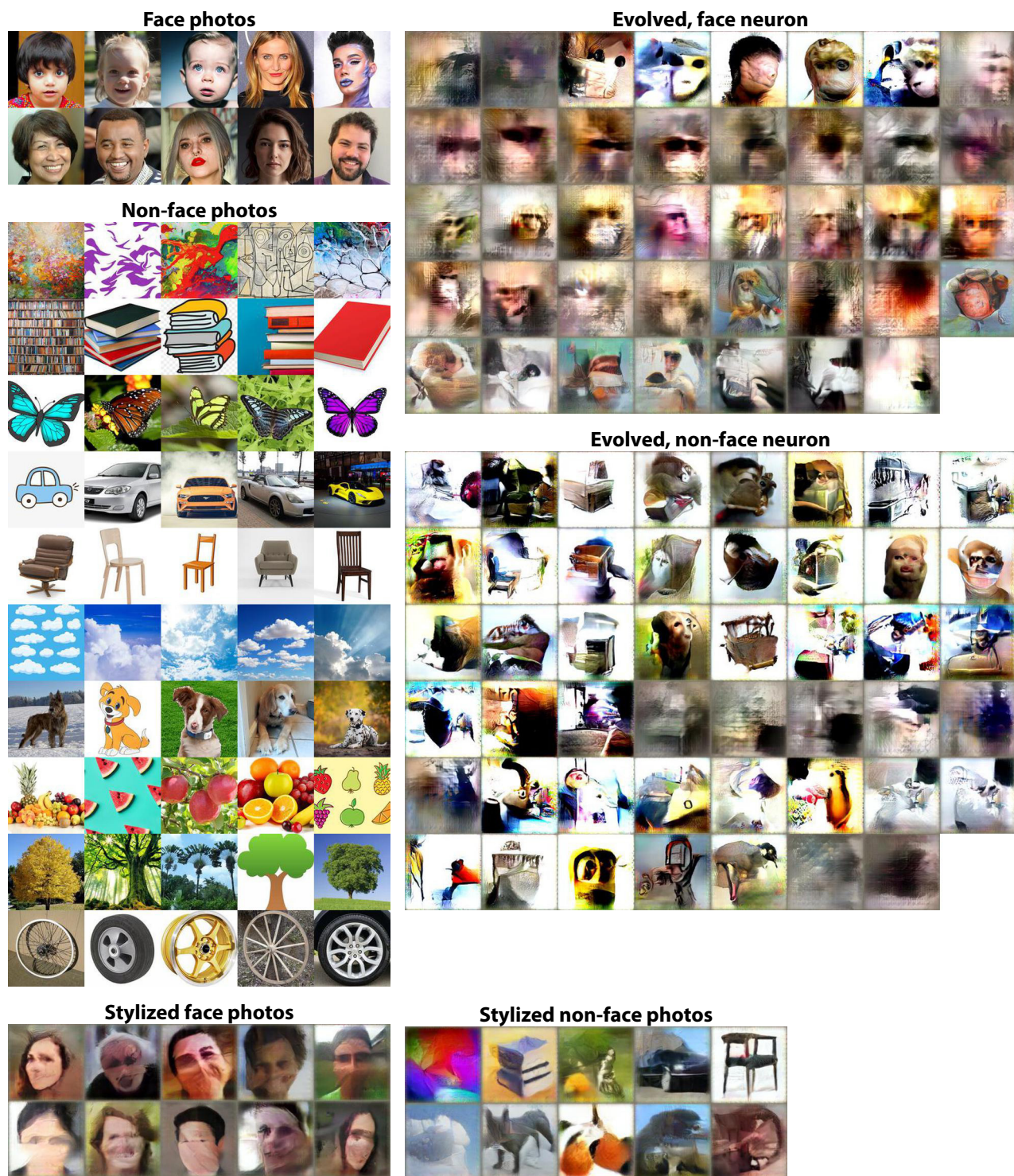
To whom correspondence should be addressed:

Gabriel Kreiman, E-mail: [gabriel.kreiman@childrens.harvard.edu](mailto:gabriel.kreiman@childrens.harvard.edu)

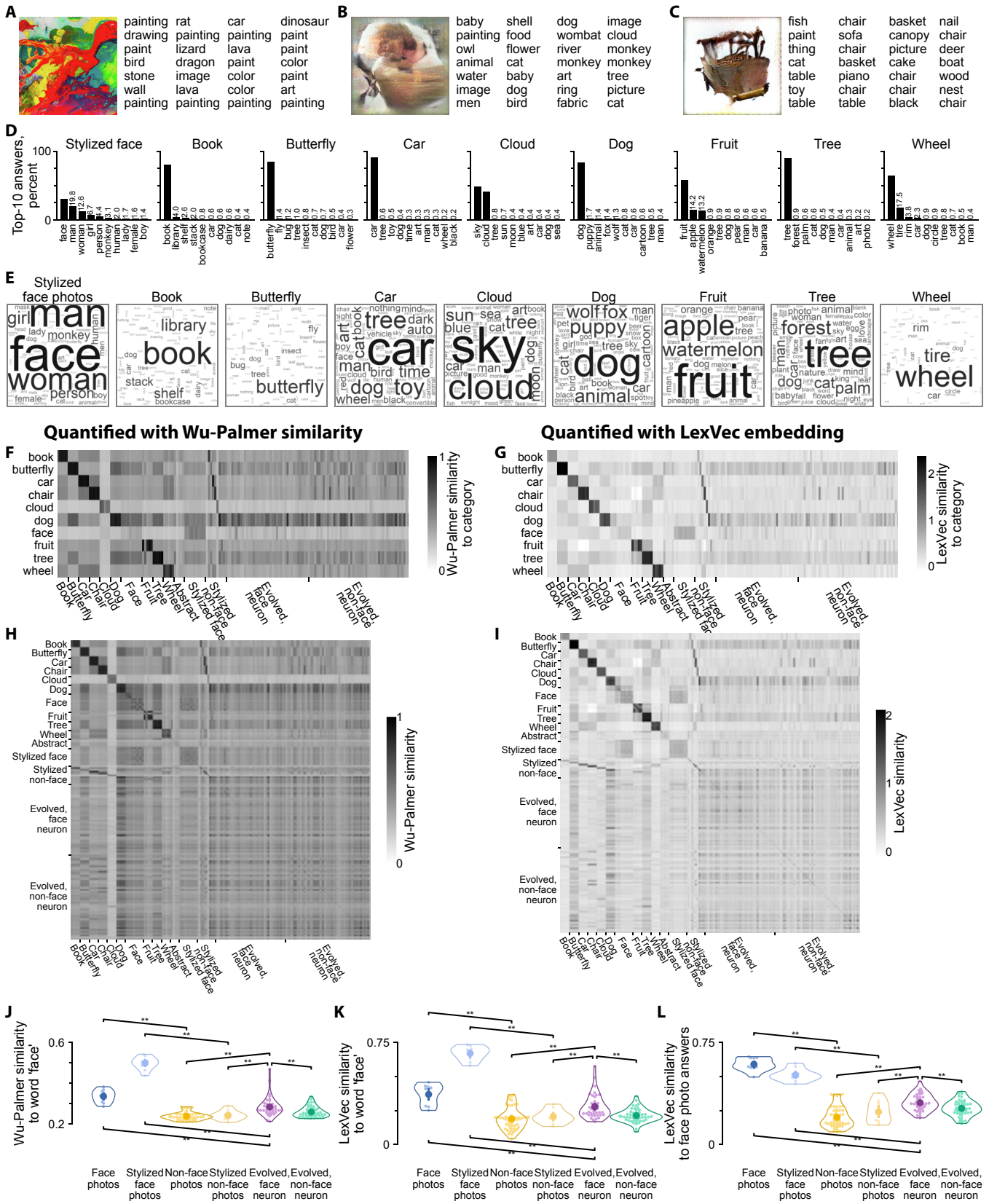
Margaret Livingstone, E-mail: [margaret\\_livingstone@hms.harvard.edu](mailto:margaret_livingstone@hms.harvard.edu)

**This PDF file includes:**

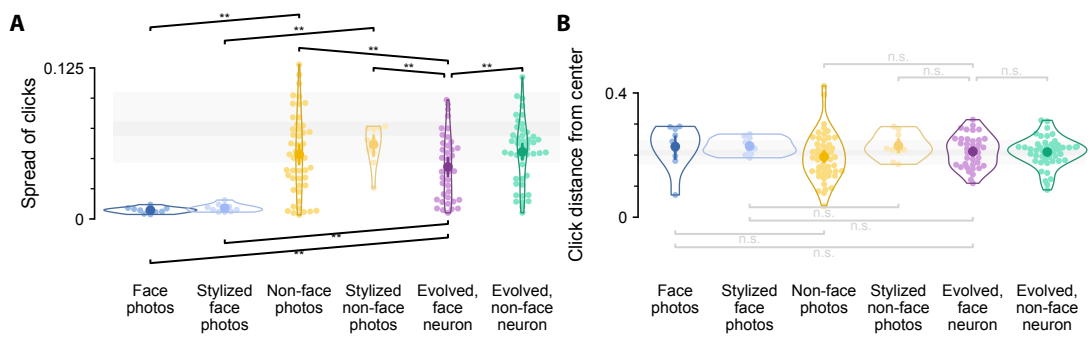
Figs. S1 to S7



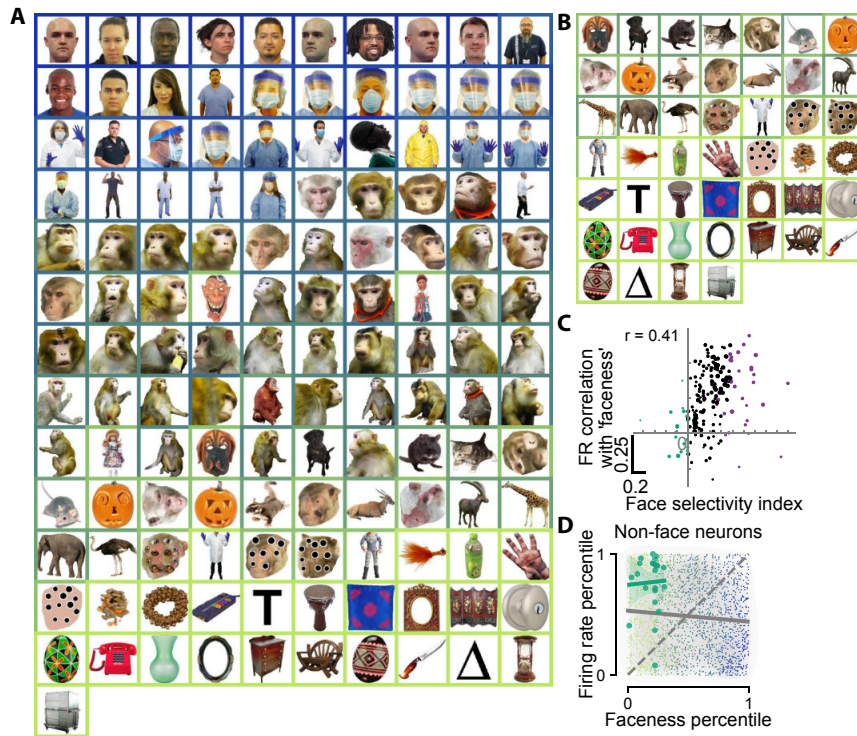
**Fig. S1. Images used in Experiments 1–6.** The set of images in Experiments 1–6 included face photos ( $n = 10$ ), non-face object images (10 categories of objects, 5 exemplars per category), stylized face images ( $n = 10$ ), and stylized non-face images ( $n = 10$ ). The main analyses focused on a set of 39 images evolved from face neurons and 47 images evolved from non-face neurons.



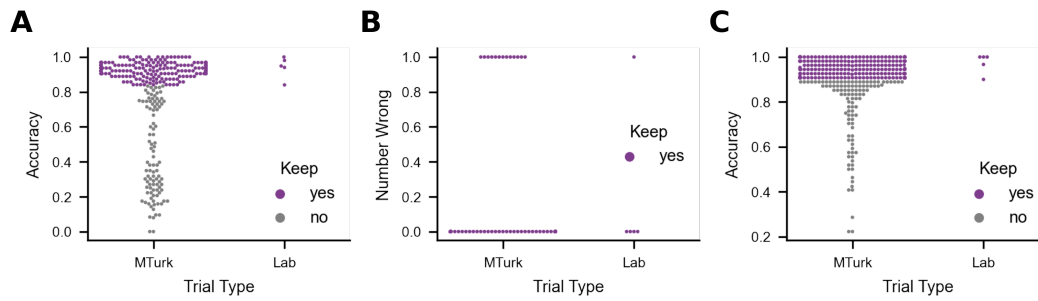
**Fig. S2. Additional analyses for Experiment 1.** **A–C.** Three example images and example responses are shown. From left to right, images are from the abstract, face neuron evolved, and non-face neuron evolved categories. **D, E.** Top-10 description frequencies (**D**) and word cloud visualization (**E**) are shown for categories not included in **Fig. 2D, E**, in the same format. **F.** Semantic (Wu-Palmer) similarity was calculated between ten category labels (y-axis) and descriptions of each image (one column). The heat map shows average similarity over descriptions of each image to each label. **H.** The heat map shows Wu-Palmer similarity between every pair of images, averaged over every pair of words between each image pair. **Fig. 2F** is a summary of the rows corresponding to 'Face' in this heat map. **J.** The swarm plot shows average Wu-Palmer similarity between descriptions of each image and the word 'face'. This is a summary of the rows corresponding to 'face' in panel **F**. Plot conventions follow those in **Fig. 2F**. **G, I, K.** Same as **F, H, J**, but using LexVec word embedding to quantify word similarity. **L.** Same as **Fig. 2F** but using LexVec similarity. **Alexandra Bardon, Will Xiao, Carlos R. Ponce, Margaret S. Livingstone, and Gabriel Kreiman**



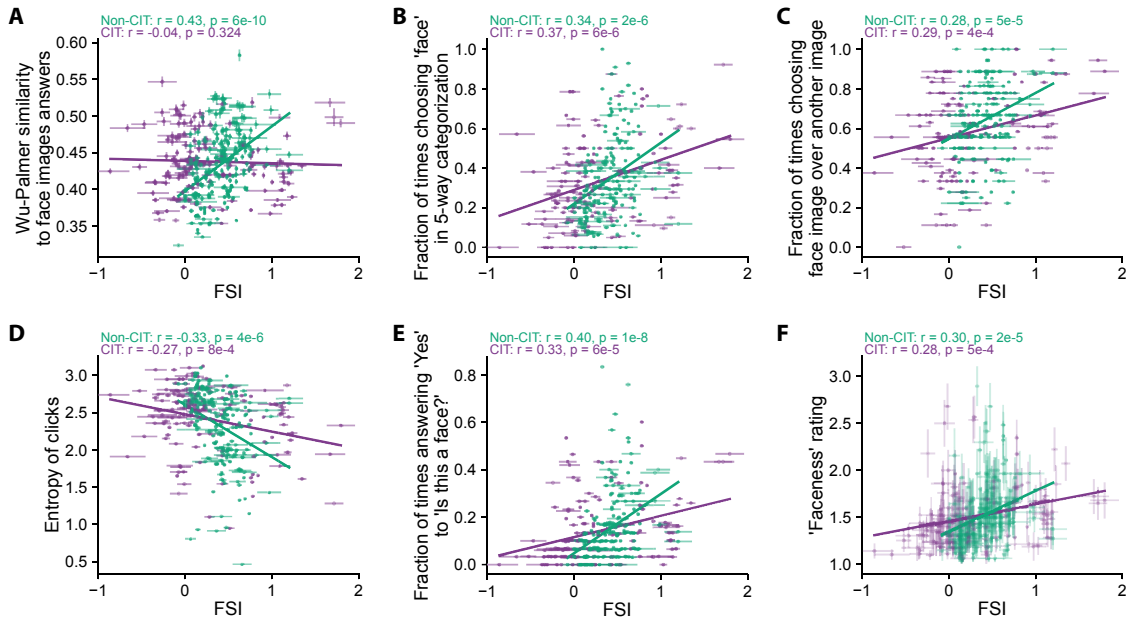
**Fig. S3. Additional measures of consistency in localizing the mouth (Experiment 4).** The swarm plots show spread of clicks (**A**) and distance from center of image (**B**). Plot conventions follow those in **Fig. 5B**



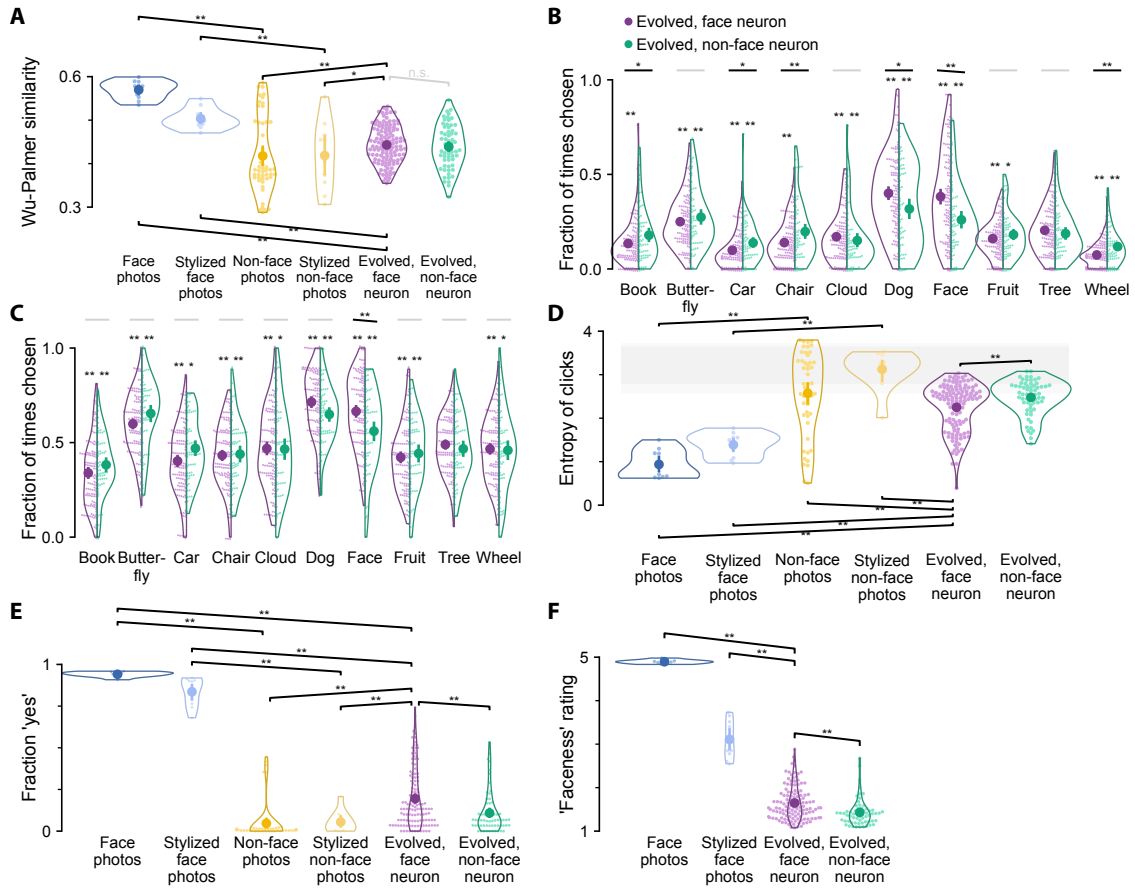
**Fig. S4. Face neuron firing rates correlated with 'faceness' rating even among non-face object images, while non-face neuron firing rates did not correlate with faceness.** **A.** Faceness ratings were collected for 131 reference images that were shown during evolution experiments. Images are sorted in reading order by their average faceness rating. Border colors correspond to the categories in **Fig. 6A**. **B.** The subset of non-face object images from **A** was used to repeat the analysis in **Fig. 6C**. **C.** The result of this analysis is shown here. **D.** The analysis in **Fig. 6E** is repeated for non-face neurons.



**Fig. S5. Comparison of subject performance in the lab versus on Mechanical Turk.** The swarm plots show response accuracy for subjects recruited on Mechanical Turk or in the lab. In-lab subjects ( $n = 5$ ) were monitored while completing experiments. Only Mechanical Turk trials for subjects with performance at or above the lowest in-lab performance were included in analyses. Color indicates whether the accuracy is above (purple) or below (gray) the minimum accuracy cutoff. Points are spread along the horizontal axis only for visualization. **A**, Performance in Experiment 2 (forced-choice five-way categorization). 'Accuracy' indicates the fraction of natural object images correctly categorized (chance is 20%). **B**, Performance in Experiment 4 (locating the mouth in an image). 'Number wrong' indicates the number of clicks outside the bounding box around the mouth for one face image. **C**, Performance in Experiment 5. 'Accuracy' indicates the fraction of 'yes' answers given for face photos (chance is 50%).



**Fig. S6. Main analyses repeated using a face-selectivity index assigned to each neuron.** Instead of labeling each neuron binarily as a face or non-face neuron, we repeated the main analyses using a continuous face-selectivity index for each neuron. Black line indicates least-squares fit. P-values are one-tailed, uncorrected. Error bars indicate S.E.M.



**Fig. S7. Main analyses repeated including neurons with imputed FSI values.** Neurons with missing or unreliable FSI estimates (based on standard deviation of bootstrap distribution over images) were assigned with the mean FSI value of other neurons, if any, recorded from the same array, then categorized as face neurons or non-face neurons using the original criteria. Main analyses for Experiments 1–6 are repeated.