nature
https://doi.org/10.1038/s41586-020-2350-5

Supplementary information

A map of object space in primate inferotemporal cortex

In the format provided by the authors and unedited

Pinglei Bao✉, Liang She, Mason McGill & Doris Y. Tsao✉

Supplementary Information for

A map of object space in primate inferotemporal cortex

Pinglei Bao[1,2,*], Liang She[1], Mason Mcgill[3], and Doris Y. Tsao[1,2,3,*,†]

1. Division of Biology and Biological Engineering, Caltech, Pasadena CA 91125.
2. Howard Hughes Medical Institute, Pasadena, CA 91125.
3. Computation and Neural Systems, Caltech, Pasadena CA 91125.

Supplementary Notes

The first two axes of object space do not depend on the precise image set or network

It might seem suspiciously serendipitous for IT to be organized according to the first two dimensions of object space, given that these first two dimensions were computed using a specific image set with a specific deep convolutional network. We confirmed these first two axes in fact do not depend strongly on the particular image set (**Extended Data Fig. 4d-f**) or network (**Extended Data Fig. 4g-j**) used to compute them. The correlations between the first two PCs computed using our original image set and using a subset of 1224 images randomly sampled from a new image database (consisting of 18,700 images of non-face objects without backgrounds and 600 face images from the FEI database, **Extended Data Fig. 2d**) were highly significant (0.92 for PC1 and 0.84 for PC2; both significant at $p < 10^{-15}$, **Extended Data Fig. 4e**). The first two PCs of object space also did not depend on the precise network used to compute them. **Extended Data Fig. 4i** shows the correlation between the first two PCs computed using AlexNet and those computed using 8 other networks (Vgg-f[1], Vgg16[2], Vgg19[2], Googlenet[3], Inceptionv3[4], Resnet101[5], Densenet201[6], and Inceptionresnetv2[7]). In each case, the correlations were highly significant ($p < 10^{-12}$). Furthermore, replotting **Fig. 4b** using PC1 and PC2 computed from 8 other networks reveals, in each case, four distinct clusters spanning PC1-PC2 space corresponding to the four anatomical networks (**Extended Data Fig. 4j**). Overall, the robustness of the first two axes of object space to the specific image set and network used to compute them suggests that they are fundamental properties of object space.

Details regarding the object space map

Early reports suggested that IT is organized into "feature columns," groups of cells ~0.5 mm in diameter sharing common feature selectivity[8,9]. The patches we report here are substantially larger. Recordings from multiple grid holes suggest each patch spans 3-4 mm (**Extended Data Fig. 8a-d**). Within single grid holes spanning just 1 mm, we found that clustering was not precise, with the extent of scatter in preferred axis in a single grid hole on the order of one quadrant of PC1-PC2 space (**Extended Data Fig. 8e, f**). Furthermore, even though cells within the IT object space map carry a rich object code consisting of many more than the two dimensions (PC1, PC2) defining the map, we did not find clustering for any dimensions beyond the first two, further suggesting that dimensions beyond the first two are distributed across the IT object map (**Extended Data Fig. 8g, h**). It is possible that finer scale organization exists; our electrode tracks were not perfectly normal to the cortical surface, and recordings from single grid holes might have spanned multiple fine columns. Dense recordings spanning each of the networks and their border regions and/or high-resolution imaging will be necessary to fully clarify whether there exists finer spatial organization within each network, and whether transitions between networks are continuous or discrete.

We found clear evidence for at least three full maps of object space in IT. Previous studies suggest the existence of six face patches in each hemisphere, with some individual variability[10]. Thus we think there may exist additional copies of the object space map in IT. Initial microstimulation experiments in monkey M1 revealed three connected patches (NML1, NML2, NM3, **Fig. 1b**, **Extended Data Fig. 1**). FMRI experiments contrasting activation to the five most- and least-preferred objects determined from single-unit electrophysiology in M1 also revealed three patches in monkeys M1-M4, with correspondence to the patches identified by microstimulation in monkey M1 (**Fig. 2c**). Subsequent fMRI experiments using stimuli from the four quadrants of object space (**Fig. 4a**) revealed four NML patches in both monkeys M3 and M4 (**Fig. 4c, d**). Based on anatomical location, NML1 targeted for electrophysiology in **Fig. 2a1** in monkeys M1 and M2 corresponds to the most posterior NML patch in monkeys M3 and M4 (**Fig. 4c** posterior**, Fig. 4d** posterior group), NML2 corresponds to the second most posterior NML patch in M3 and M4 (not shown in **Fig. 4c, Fig. 4d** not outlined), and NML3 corresponds to the third most posterior NML patch in M3 and M4 (**Fig. 4c** middle**, Fig. 4d** middle group).

Decoding analysis

To quantify the object information available in the map of object space formed by the four networks, we analyzed object decoding accuracy as a function of number of distractor objects, number of neurons used to build the decoder, and number of object feature dimensions. **Extended Data Fig. 11d** shows decoding accuracy as a function of number of distractor objects (see Methods). We compared the actual object feature vectors of a subset of images to the reconstructed feature vector for one image ("target") using a Euclidean distance metric. If the actual object feature vector with the smallest distance to the reconstructed object feature vector portrayed the target, the decoding was considered correct. For 39 distractors, decoding accuracy was ~ 0.65 when cells across four networks were combined. Decoding accuracy increased as a function of number of cells within each network (**Extended Data Fig. 11e**). Object decoding accuracy increased with addition of new object space dimensions, up to 10 (**Extended Data Fig. 11f**; note this is just a lower bound, and the dimensionality would likely increase even further with use of other stimulus sets).

Intuitively, neurons in certain networks should be more discriminative of certain stimuli. To quantify this, for each object and each network, we computed a "specialization index" SI_{ij} that measures how much better decoding accuracy for object i computed from activity in network j is compared to decoding accuracy for object i computed across all other networks, using the same number of neurons in each case (see Methods). Distributions of SI_{ij} across 1224 objects in the NML and body networks revealed many objects with values significantly greater than zero (t-test two-tailed $p<0.01$, **Extended Data Fig. 11i**), indicating that they have a specialized representation by the corresponding network. A predicted consequence is that perturbation of specific networks should differentially affect object recognition behavior. This is supported by a recent study showing that inactivating millimeter-scale IT subregions results in selective object recognition deficits[11].

Table 1

ID	*L/R*	*Face*	*Body*	*NML*	*Stubby*	*Color*	*Disparity*	*Scene*	*Covered IT*	*Whole IT*	*Covered Percentage*
M1	L	104	102	82	N/A	45	36	30	340	618	55%
M1	R	90	78	76	N/A	35	60	30	320	603	51%
M2	L	90	82	95	82	40	60	40	320	570	56%
M2	R	85	95	91	93	40	50	45	315	560	56%
M3	L	84	49	66	85	30	71	N/A	299	599	50%
M3	R	102	67	99	85	22	35	N/A	307	533	58%
M4	L	73	60	78	81	27	45	N/A	241	510	47%
M4	R	110	93	60	100	45	50	N/A	306	505	59%
Avg		90	76	84	85	34	51	36	306	570	53%

Table 1. Coverage of IT cortex by the object-topic map. The table lists the cortical area in mm^2 for seven different networks identified by seven localizers for the four macaque subjects used in this study. IT was defined as the aggregation of TE and TEO (as defined by [12,13]).

	Prediction	Justification	Experimental evidence
Anatomic layout	1. IT should contain a network representing stubby objects.	If IT contains a coarse map of object space, then distinct quadrants of object space should map to distinct networks.	Fig. 2d, Fig. 4b, c
	2. The spatial layout of patches should follow the arrangement of preferred stimulus clusters in object space.	If IT contains a coarse map of object space, then the topography of object space should be conserved.	Fig. 4e-g
	3. The object space model should parsimoniously explain existing experimental findings about large-scale IT organization.	The object space model is a comprehensive, computable model.	Extended Data Fig. 9
Coding principles	4. Coding principles of the face network should generalize to other networks in IT.	In the object space model of IT, face patches are not unique, but simply responsible for representing one part of object space.	Fig. 2a-d (consistent visual selectivity) Fig. 3 (increasing invariance, axis model)
	5. IT cortex is organized according to the first two axes of object space rather than low-level features or image organization.	The object space of AlexNet fc6 is not based on low-level features.	Extended Data Fig. 6c, e
	6. IT cortex is organized according to shape-based object space dimensions rather than high-level semantic dimensions.	The object space of AlexNet fc6 is not based on semantic identity.	Extended Data Fig. 6d, Extended Data Fig. 10
	7. The first two axes of object space should not depend critically on the precise deep network or image set used to compute them.	Initial support for the idea that IT is laid out according to object PC1-PC2 space was based on axes computed using a specific set of 1224 images run through a specific deep network, AlexNet. Unless we were exceptionally lucky, these axes should generalize.	Extended Data Fig. 4d-j
Behavior	8. Different regions of IT should make different contributions to object recognition behavior.	Many stimuli are better discriminated by one network compared to others, depending on their first two components in object space.	Extended Data Fig. 11i Rajalingham et al., Neuron, 2019

Table 2. A table summarizing 8 predictions that follow from the hypothesis that IT contains a coarse map of object space, together with justifications and experimental evidence.

1 Chatfield, K., Simonyan, K., Vedaldi, A. & Zisserman, A. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531* (2014).

2 Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

3 Szegedy, C. *et al.* in *Proceedings of the IEEE conference on computer vision and pattern recognition.* 1-9.

4 Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. in *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2818-2826.

5 He, K., Zhang, X., Ren, S. & Sun, J. in *Proceedings of the IEEE conference on computer vision and pattern recognition.* 770-778.

6 Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2261-2269 (IEEE).

7 Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. A. in *AAAI.* 12.

8 Wang, G., Tanaka, K. & Tanifuji, M. Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* **272**, 1665-1668 (1996).

9 Fujita, I., Tanaka, K., Ito, M. & Cheng, K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* **360**, 343 (1992).

10 Tsao, D. Y., Moeller, S. & Freiwald, W. A. Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences* **105**, 19514-19519 (2008).

11 Rajalingham, R. & DiCarlo, J. J. Reversible inactivation of different millimeter-scale regions of primate IT results in different patterns of core object recognition deficits. *Neuron* **102**, 493-505. e495 (2019).

12 Reveley, C. *et al.* Three-dimensional digital template atlas of the macaque brain. *Cerebral Cortex* **27**, 4463-4477 (2016).

13 Saleem, K. S. & Logothetis, N. K. *A combined MRI and histology atlas of the rhesus monkey brain in stereotaxic coordinates*. (Academic Press, 2012).