nature neuroscience

Only some spatial patterns of fMRI response are read out in task performance

Mark A Williams[1,2], Sabin Dang[1] & Nancy G Kanwisher[1,3,4]

Classification methods show that the spatial pattern of a functional magnetic resonance imaging response across the cortex contains category information, but whether such patterns are used, or 'read out', in behavioral performance remains untested. We show that although the spatial pattern in both the retinotopic and lateral occipital cortex (LOC) in humans contains category information, only in the LOC is the pattern stronger for correct than for incorrect trials. Thus, some, but not all, spatial patterns are read out during task performance.

Traditional functional magnetic resonance imaging (fMRI) data analyses apply statistical tests separately to each voxel, asking, for example, whether the response of that voxel is significantly higher in condition A than in condition B. Recently, however, new multivariate classification methods have become very popular[1–7]. These new analyses use the spatial pattern of an fMRI response across the cortex to determine what type of information is present in each region of the cortex. Previous studies have used these analyses to predict whether subjects were viewing chairs versus bottles[1], attending to one stimulus orientation versus another[3] and telling a lie versus a truth[5]. With all the excitement about classification methods, however, a fundamental question has remained unanswered: is the information carried in these spatial patterns actually used or is it epiphenomenal (that is, not causally related to task performance)?

Here we propose a method for addressing this question that is based on the following idea: if task performance is determined by the spatial pattern of the fMRI response in a given brain region, then discriminative information in that spatial pattern should be higher on correct trials than on incorrect trials. As a proof of concept of this method, we used it to determine which cortical regions were directly involved in a visual shape discrimination task.

Previous studies have implicated cortical regions of the ventral visual pathway in visual recognition by binning fMRI data according to behavioral response and showing that neural activity in some of these regions is higher on correct than on incorrect trials[8–10]. Here, we combine this method of binning fMRI data by behavioral response with pattern classification methods to ask not just which regions have a mean response that is related to behavior, but which regions contain information that is related to behavior (**Supplementary Methods** and **References** online).

Six subjects were scanned while viewing three categories of novel objects ('spikies', 'smoothies' and 'cubies'[11]; **Supplementary Fig. 1** online). Informed consent was obtained according to the procedures approved by the Massachusetts Institute of Technology and the Massachusetts General Hospitals Committee on the Use of Humans as Experimental Subjects. Single objects were presented briefly and masked, and subjects indicated which category each object belonged to (**Supplementary Fig. 2** online). The critical regions of interest (ROIs) identified in an independent localizer scan were the LOC and foveal representation in retinotopic cortex. Spatial patterns were extracted from each half of the data (odd versus even runs) separately for each combination of ROI, stimulus category and response (correct versus incorrect). Within each ROI we then computed the correlation between the spatial patterns on odd and even runs from the same stimulus category (for example, spiky-spiky) versus different stimulus categories (for example, spiky-smoothie); this analysis was performed separately on the data from trials in which the subject performed correctly versus incorrectly in categorizing the object. A higher correlation in a given ROI for same-category spatial patterns than for different-category spatial patterns indicated the

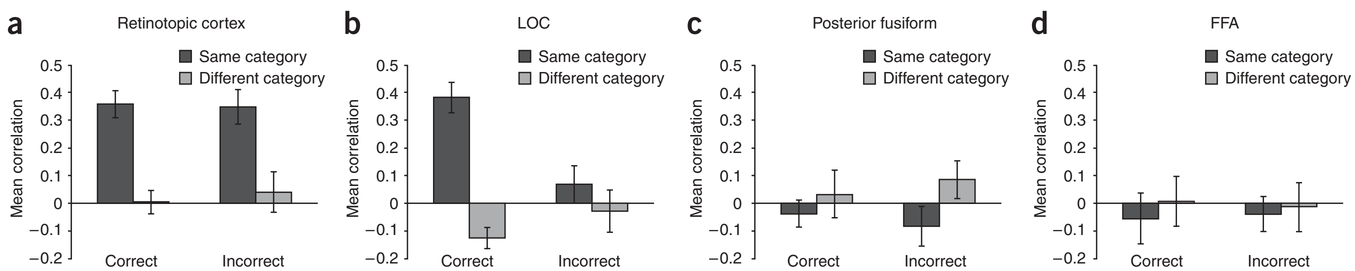

Figure 1 Mean correlations (± 1 s.e.m.) same and different categories for correct and incorrect trials in four ROIs. (**a**) Foveal representation in retinotopic cortex (central object presentation > fixation). (**b**) Lateral occipital complex (defined by a comparison of objects > scrambled objects). (**c**) Posterior fusiform gyrus (objects > scrambled objects). (**d**) FFA (faces > objects).

[1]McGovern Institute for Brain Research, 77 Massachusetts Avenue, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Macquarie Centre for Cognitive Science, Macquarie University, Sydney, NSW 2109, Australia. [3]Brain & Cognitive Sciences, 77 Massachusetts Avenue, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Athinoula A. Martinos Center for Biomedical Imaging, 149 Thirteenth Street, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. Correspondence should be addressed to M.A.W. (markow@mit.edu).

Received 27 December 2006; accepted 27 March 2007; published online 7 May 2007; doi:10.1038/nn1900

presence of discriminative information in the spatial patterns in that ROI.

In the retinotopic ROI, same-category correlations were higher than different-category correlations ($F_{1,5} = 32.58$, $P < 0.005$), and this was true to the same degree for both correct trials ($t_5 = 5.71$, $P < 0.01$) and incorrect trials ($t_5 = 3.33$, $P < 0.05$; there was no interaction of same/different by correct/incorrect, $F_{1,5} = 1.50$, $P = 0.28$; **Fig. 1a**). In contrast, correlations across spatial patterns in the LOC were higher for same-category than for different-category pairs, but this was true only for correct trials ($t_5 = 3.40$, $P < 0.05$) and not for incorrect trials ($t_5 = 1.23$, $P = 0.26$; this interaction was significant, $F_{1,5} = 8.31$, $P < 0.05$; **Fig. 1b**). These findings show that although spatial patterns in the retinotopic ROI and the LOC both contain category information, this information is directly related to behavior only in the LOC and not in the retinotopic cortex.

Analyses of other brain regions (posterior fusiform and fusiform face area, FFA) did not show discriminative information in the spatial pattern of response either when we analyzed all data or when we analyzed only correct trials. However, we cannot rule out the possibility that discriminative information may exist in these or other brain regions, perhaps at a finer grain.

Further analyses ruled out a number of alternative accounts for our findings. First, it is the spatial pattern, not mean response across voxels, that contains category information and that is related to behavior in LOC; there was no interaction of stimulus category × correct/incorrect ($F_{1,5} = 1.26$, $P = 0.31$) and there was no main effect of category ($F_{1,5} = 2.58$, $P = 0.17$) or accuracy ($F_{1,5} = 0.41$, $P = 0.55$) (**Supplementary Fig. 3** online) on mean response. Second, the lower discriminative information for incorrect than for correct trials in LOC was not due to the lower number of incorrect than correct trials, because a reanalysis using matched trial numbers (after removal of a random subset of correct trials) replicated the significant discrimination between categories on correct ($t_5 = 5.39$, $P < 0.05$), but not incorrect ($t_5 = 1.23$, $P = 0.26$), trials (**Supplementary Fig. 4** online; this interaction was significant, $F_{1,5} = 6.39$, $P < 0.05$). Third, the presence of discriminative information in the LOC but not the FFA was not due to the smaller number of voxels in the latter, as the same results were obtained (**Supplementary Fig. 5** online) when the number of voxels in the LOC was equated with those in the FFA (by increasing the statistical threshold for inclusion of voxels in the LOC).

In sum, pattern information in some cortical regions (for example, LOC) is related to behavior (that is, is stronger on correct than on incorrect trials), whereas pattern information in other regions (for example, retinotopic cortex) is not. How are we to interpret these data? The retinotopic cortex is presumably critically involved in the performance of this task. Yet, notably, on the trials in which subjects reported the wrong stimulus category, correct stimulus information is nonetheless present in the retinotopic cortex. What these data suggest is that the information in the LOC, but not in the retinotopic cortex, seems to be read out when subjects decide which shape category they saw. This read out hypothesis predicts that at the moment when the subjects decide which shape they saw, altering the pattern information (for example, with transcranial magnetic stimulation or microstimulation[12,13]) in the LOC would disrupt performance, but altering neural representations in V1 would not[14].

The method presented here provides a way to distinguish cortical regions where detectable pattern information is closely associated with behavioral performance, making these regions candidates for the locus of read-out, from those where the detectable pattern information is not associated with behavior, and hence not read out directly during task performance (see also ref. 15). The data presented here serve as a proof of concept of our method and underline the fact that the neural representation of an object that underlies performance need not include all of the cortical regions that contain information about that object. It will be important in future work not only to look for discriminative information in the cortex, and to characterize the different kinds of information that exist in different cortical regions, but to test what discriminative information is used. Future research might also fruitfully explore how read-out of pattern information changes with task requirements and perceptual experience.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We thank C. Baker, D. Dilks, H. Op De Beeck, J. Haushofer, A. Rich, R. Saxe, B. Schwarzlose, E. Vul and U. Zohary for their comments and suggestions on this work. M.A.W. was supported by a grant from the Australian National Health and Medical Research Council (C.J. Martin Fellowship). This work was supported by a grant from US National Institutes of Health to N.G.K (grant 13455).

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at http://www.nature.com/natureneuroscience
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Haxby, J.V. *et al. Science* **293**, 2425–2430 (2001).
2. Cox, D.D. & Savoy, R.L. *Neuroimage* **19**, 261–270 (2003).
3. Kamitani, Y. & Tong, F. *Nat. Neurosci.* **8**, 679–685 (2005).
4. Kamitani, Y. & Tong, F. *Curr. Biol.* **16**, 1096–1102 (2006).
5. Langleben, D.D. *et al. Hum. Brain Mapp.* **26**, 262–272 (2005).
6. Haynes, J.P. & Rees, G. *Nat. Neurosci.* **8**, 686–691 (2005).
7. Kriegeskorte, N., Goebel, R. & Bandettini, P.A. *Proc. Natl. Acad. Sci. USA* **103**, 3863–3868 (2006).
8. Bar, M. *et al. Neuron* **29**, 529–535 (2001).
9. Grill-Spector, K., Kushnir, T., Hendler, T. & Malach, R. *Nat. Neurosci.* **3**, 837–843 (2000).
10. Grill-Spector, K., Knouf, N. & Kanwisher, N. *Nat. Neurosci.* **7**, 555–562 (2004).
11. Op de Beeck, H.D., Baker, C.I., DiCarlo, J.J. & Kanwisher, N.G. *J. Neurosci.* **26**, 13025–13036 (2006).
12. Afraz, S.R., Kiani, R. & Esteky, H. *Nature* **442**, 692–695 (2006).
13. Chambers, C.D. & Mattingley, J.B. *Trends Cogn. Sci.* **9**, 542–550 (2005).
14. Cohen, M.R. & Newsome, W.T. *Curr. Opin. Neurobiol.* **14**, 169–177 (2004).
15. Hampton, A.N. & O'Dohery, J.P. *Proc. Natl. Acad. Sci. USA* **104**, 1377–1382 (2007).