Does the fusiform face area contain subregions highly selective for nonfaces?

To the editor:
Grill-Spector and colleagues[1] used high-resolution fMRI (1 mm isotropic voxels) to argue that the fusiform face area (FFA) is composed not only of voxels highly selective for faces (as previously argued), but also of voxels that are just as selective for nonface stimuli (cars, animals and sculptures). We show here that the authors' analyses are subject to a selection bias that invalidates their claim of strong selectivity for nonfaces.

To quantify the selectivity of voxels within the FFA at high resolution, Grill-Spector and colleagues first determined the preferred category for each voxel (based only on magnitude of response, not on any statistical comparison), and then averaged across voxels with the same preference (using the same data used to determine preference) to obtain a selectivity profile for voxels preferring faces, cars, animals and sculptures. The resulting plots (**Fig. 1a**, reproduced from Fig. 4 in Grill-Spector *et al.*) suggest the existence of voxels strongly selective for nonfaces. The problem with this analysis is that it is not independent—the same data were used to select voxels and to produce the selectivity profiles. Even if the data consisted entirely of noise, voxels selected because they have a higher response to one category than to any of the others will show apparent selectivity for that category when you average across them. Independent tests are needed to show that the apparently higher response of voxels to a given category than to each of the other categories is statistically significant . While Grill-Spector and colleagues reported strong reliability of their selectivity index, this analysis is flawed (see **Supplementary Note 1**). One standard way to demonstrate reliability of selective responses is cross validation: one data set is used to define voxels based on a certain functional property, and an independent data set is used to test whether that functional property is reproducible.

Using exactly the same stimuli and experimental design, we replicated

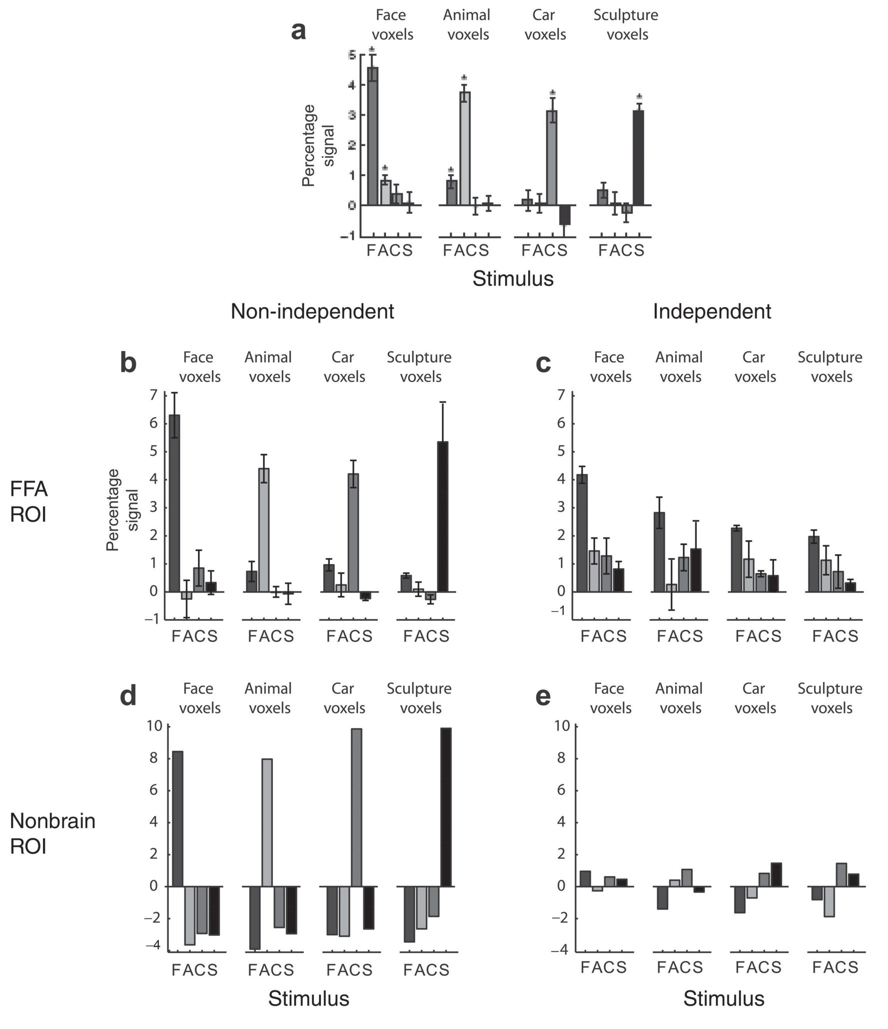

Figure 1 Comparison of results reproduced from Grill-Spector *et al.* with non-independent and independent analyses of data from FFA and non-brain regions of interest. (**a**) Data from Grill-Spector et al. (**b**,**c**) Although the non-independent analysis in the FFA ROI qualitatively replicates the results of Grill-Spector *et al.*, cross-validation analysis shows that only in the case of face voxels is this preference maintained in an independent data set (n = 3; error bars, s.e.m.). (**d**,**e**) Example data from the non-brain region of interest in one participant. Non-independent analysis qualitatively replicates the results of Grill-Spector *et al.*

Experiment 1 from Grill-Spector *et al.* (we thank K.G.-S. for providing us with her stimuli), scanning three participants at the same high resolution (1 mm isotropic voxels) (see **Supplementary Note 2**). We then investigated the responses of voxels within the FFA region using both the non-independent analysis of Grill-Spector and colleagues and an independent cross-validation analysis (see **Supplementary Note 3**).

Using the non-independent analysis method, we replicated the results of Grill-Spector and colleagues (**Fig. 1b**), showing voxel populations apparently strongly selective for animals, cars and sculptures in addition to faces. Critically, however, this result does not survive a cross-validation test, in which the selection and test data are independent. Instead, only the face voxels maintain their preference in the independent test (**Fig. 1c**). Thus, the apparent selectivity for nonface categories in the non-independent analysis must have reflected random variation, not replicable selectivity. The bias in the non-independent analysis method can be seen most clearly when this method is used to analyze a region of interest that necessarily includes only noise because it is outside the brain (**Fig. 1d**,**e**): again, this method produces apparently selective voxel populations for all categories. Because their analysis method yields similar results when analyzing noise, the findings reported by Grill-Spector and colleagues provide no evidence for strong selectivity for nonfaces (or even faces) on the fusiform gyrus.

Grill-Spector and colleagues are correct that the apparent functional homogeneity of a given cortical region based on scans performed at one imaging resolution can be trumped by higher-resolution scans that demonstrate functional heterogeneity at a finer grain. However, they have not yet demonstrated that the FFA contains any voxels strongly selective for animals, cars or sculptures.

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

Chris I Baker[1], Tyler L Hutchison[2] & Nancy Kanwisher[2]

[1]Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, 10 Center Drive, Building 10, Room 4C-104, MSC1366, Bethesda, Maryland 20892, USA. [2]McGovern Institute for Brain Research and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.
e-mail: bakerchris@mail.nih.gov

1. Grill-Spector, K., Sayres, R. & Ress, D. *Nat. Neurosci.* **9**, 1177–1185 (2006).

Measuring selectivity in fMRI data

To the editor:
Using high-resolution functional magnetic resonance imaging (fMRI), Grill-Spector and colleagues[1] recently reported that many voxels in the fusiform face area (FFA) gave a numerically higher response to nonface stimuli than to faces. Importantly, the response to nonface objects appeared to show extremely high category selectivity. As such, these findings pose a serious challenge to current models of object representation in the ventral occipitotemporal cortex[2]. The formula used by Grill-Spector to calculate selectivity, however, vastly overestimates actual voxel selectivity.

Grill-Spector calculated voxel selectivity using the following formula: Selectivity = [Preferred – Nonpreferred] / [Preferred + |Nonpreferred|]. In this formula, "Preferred indicates the amplitude of the category that yielded the maximal response and Nonpreferred indicates the average amplitude of other categories" (p. 1184), with a selectivity value of 1 indicating maximum selectivity, and 0 indicating that a voxel has no preference for any of the stimulus categories tested. This formula is similar to the standard formula used in many electrophysiological studies, with the exception that the absolute value of the average nonpreferred response was used in the denominator. Unlike spike rate data, however, negative values are commonly observed in fMRI, and this has significant consequences when used with this formula. Consider the following. Suppose that the responses in a particular voxel were 0.11 to one category and 0.10, 0.08 and –0.20 to the other categories. Using Grill-Spector's formula, this voxel's selectivity equals 1.0, even though the response difference between the preferred and nearest non-preferred category equals only ~ 3% of the voxel's total response range (0.11 to –0.20). Clearly, this voxel is not highly selective for the preferred category. In fact, using this formula, anytime

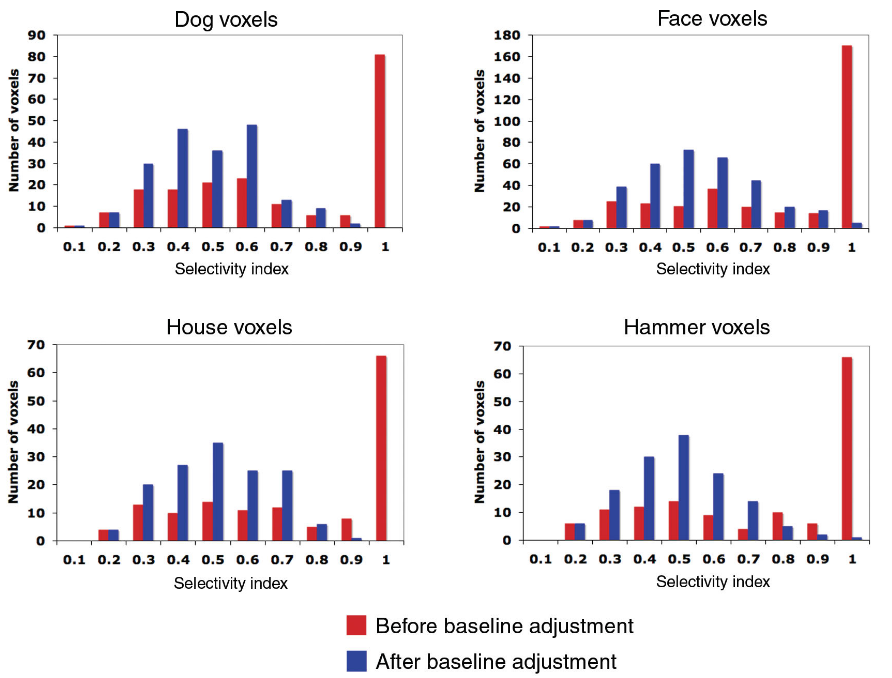

Figure 1 Histograms showing the selectivity index of FFA voxels with respect to each of four object categories, averaged across four subjects. Red bars show selectivity values before baseline adjustment. Blue bars show selectivity values for the same data after adjusting the signal baseline. (This analysis is an attempt to replicate the methods of Grill-Spector *et al.* with a corrected selectivity formula. No attempt was made to correct for the selection bias discussed by Baker *et al.* in the accompanying correspondence, which likely explains why the mean selectivity is comparable for all categories.)