Supporting Information

Supporting Information Corrected August 29, 2013
McGugin et al. 10.1073/pnas.1116333109

SI Text

Choice of Baseline for Correlations with Behavioral Expertise. In prior studies looking at individual differences correlated with the fusiform face area (FFA) activity (e.g., refs. 1 and 2) a "high-level" baseline is used. This use is because the goal is to correlate behavioral performance to variance in neural activity that is specific to response for cars (not to variance that reflects how much an individual subject's visual cortex activates to any visual image). For example, here, if one looks at the correlation across all subjects for the brain activity between the neural response for any two categories (e.g., faces and cars), we tend to observe very strong correlations in many areas when using a fixation baseline (r values around 0.8), and then still relatively strong correlations when using a scrambled baseline (r values around 0.5), but much lower values when using a high-level baseline like animals (r values around 0.2). Although other studies have found expertise effects for cars using as a baseline the response to faces or birds, here we used animals as the baseline category because the response to animals was not expected to vary as a function of expertise and indeed, the response to animals relative to scrambled images was almost never correlated with car expertise (the exception being the left FFA face-sensitive voxels, $r = 0.54$).

Traditionally, correlations between expertise effects and neural activity have been computed on percent signal change (PSC) indices, but because studies of selectivity in the FFA in man and monkey have also used a normalized measure of neural sensitivity (3, 4), we included such a measure (d_a) in which the response to the preferred category is compared with that for nonpreferred categories, normalized by the SD of activity for preferred and nonpreferred, to demonstrate the generality of expertise effects across different measures of category preference. Another advantage of using d_a is that it provides a means to compare the results of the main experiment with those of Exp. 2, in which animals were not presented and could not be used a baseline.

Outlier Detection for Correlation Analyses. For all reported correlations between behavioral expertise and a neural measure of sensitivity, we tested for bivariate outliers by computing externally studentized residuals for each correlation. Residuals greater than 3.5 or less than −3.5 were considered outliers and the individual points were removed from the analysis. One outlier was observed in 20% of the correlations tested (not systematically the same subject), and no correlation had more than one.

Time-Series Signal-to-Noise Ratio. One concern in the comparison of activation from regions of interest (ROIs) and rings of varying expanses is the time-series signal-to-noise ration (tSNR) of outer rings relative to inner ROIs. Because areas of cortex around the peak may show reduced tSNR, it is important to demonstrate that the absence of an expertise effect in the 300- to 200-mm^2 ring is not simply because of a decrease in signal. The Scrambled > Fixation signal is compared across two ROIs (25 mm^2 and 100 mm^2) and two ROI-rings (200–100 mm^2 and 300–200 mm^2), demonstrating consistent signal across varying expanses of cortex surrounding the peak. The tSNR was above 40 in all ROIs [mean, SEM: 25 mm^2 = 46.9 (5.5), 100 mm^2 = 48.4 (6.1), 200–100 mm^2 = 45.7 (5.3), 300–200 mm^2 = 45.5 (5.0)] and there was no drop in signal in the 300- to 200-mm^2 ROI-ring compared with any of the other ROIs (vs. 200–100 mm^2, $P = 0.20$; vs. 100 mm^2, $P = 0.60$ and vs. 25 mm^2, $P = 0.51$), showing that the spatially restricted expertise effect cannot be caused by a decline in signal quality in the largest ring.

SI Methods

Experiment 2. ***Subjects.*** Twenty-six healthy adults (six females, aged 22–41 y with an average of 27 y) participated for monetary compensation. Informed written consent was obtained from each participant in accordance with guidelines of the institutional review board of Vanderbilt University and Vanderbilt University Medical Center. All participants had normal or corrected-to-normal visual acuity.

Scanning. All imaging was performed on a Philips Medical Systems 7-Tesla (7T) human magnetic resonance scanner at the Institute of Imaging Science at Vanderbilt University Medical Center (Nashville, TN).

High-resolution anatomical scan. High-resolution (HR) T1-weighted anatomical volumes were acquired using a 3D TFE (Turbo Field Echo) acquisition sequence with sensitivity encoding (SENSE) [field-of-view (FOV) = 246 mm, minimum TE, TR = 3.152 ms, matrix size = 352 × 352] to obtain 249 slices of 0.7-mm^3 isometric voxels. HR anatomical images were used to align sets of functional data, for volume rendering (including gray matter–white matter segmentation for the purposes of inflating and flattening of the cortical surface), and for visualization of functional data.

Standard-resolution functional scan. The experimental sequence began with a single-run standard-resolution (SR) functional localizer for real-time localization of FFA and optimal positioning of HR slices. We acquired 30 SR slices (2.3 × 2.3 × 2.5 mm) oriented in the coronal plane. The blood-oxygen level-dependent–based signals were collected using a fast T2*-sensitive radio-frequency-spoiled 3D PRESTO sequence (FOV = 211.2 mm, TE = 22 ms, TR = 21.93 ms, volume repetition time = 2,500 ms, flip angle = 78°, matrix size = 96 × 96). Slice prescriptions that incorporated face-responsive voxels (localized on an individual basis using real-time linear regression analysis during the SR functional scan) and also included the fusiform gyrus (FG) were selected for subsequent HR runs.

HR functional scans. Immediately following the SR scan, we acquired 24 HR oblique slices. We used a radio frequency-spoiled 3D FFE acquisition sequence with SENSE (FOV = 160 mm, TE = 21 ms, TR = 32.26 ms, volume repetition time = 4,000 ms, flip angle = 78°, matrix size = 128 × 128) to obtain 1.25-mm^3 isotropic voxels. Three functional scans with these parameters were acquired in each participant dedicated to localization of face, car and sofa-preferring voxels.

Functional MRI Display, Stimuli, and Task. All images were presented on an AppleMacintosh computer using Matlab (MathWorks) with Psychophysics Toolbox extension (5, 6). We used 72 grayscale images (36 faces, 36 objects) to localize face- and object-selective regions in the SR run. The three HR face-car-sofa runs used another 102 grayscale images each of faces, cars (nonfrontal views) and sofas. Scrambled images were created on-line during the scan by parsing a randomly selected image into 64 equally sized squares, then rearranging these squares to form a novel display of scrambled pieces. All stimuli used in the SR face-object run and HR face-car-sofa runs were presented in the center of the screen and resized on each trial to subtend a visual angle of ∼12.6, ∼15.1, or ∼17.6. The stimuli were presented at different sizes to discourage the use of low-level visual properties to perform the task.

SR Localizer run. The localizer scan used a one-back detection task with 14 blocks of alternating faces and common objects (20 images shown for 1 s) with a 2.5-s fixation at the beginning and end. This scan assisted identification of face-selective voxels (Face > Object) in the inferior-temporal and occipital cortices, thereby guiding

the prescription of HR slices. Accuracy did not differ for Face and Object blocks: (hit rate, false-alarm rate) Face (0.89, 0.01), Object (0.88, 0.01).

HR experimental runs. Immediately following real-time alignment of the HR slices based on SR data, participants completed three additional runs using a 1-back detection task. Each run contained 14 blocks (four each of faces, cars and sofas, and two scrambled) of 16-s duration each (16 images sequentially presented for 1 s each), with 4-s fixation at the beginning and end. A one-way ANOVA revealed a main effect of category ($F_{1,25} = 105.74, P > 0.0001$), carried by significantly poorer detection during Scrambled blocks relative to the intact-object blocks, but no difference between Faces, Cars and Sofas (Scheffé test pairwise comparisons, $P < 0.0001$): (hit rate, false-alarm rate) Face (0.93, 0.007), Car (0.93, 0.005), Sofa (0.91, 0.01), Scrambled (0. 24, 0.03).

Behavioral Expertise Measure and Stimuli. A behavioral task outside the scanner used a separate set of grayscale images of 56 cars and planes. Subjects performed 12 blocks of 28 sequential matching trials per category. On each trial, the first stimulus appeared for 1,000 ms followed by a 500-ms mask. A second stimulus appeared and remained visible until a same or different response or 5,000 ms. Subjects judged whether the two images showed cars or planes from the same make and model regardless of year. An expertise sensitivity score was calculated for cars (car d', range 0.52–3.17) and planes (plane d', range 0.80–3.13) for each subject.

Data Analysis. The data analysis in Exp. 2 was identical to that of Exp. 1 (see *Methods*), except all correlations with car expertise in Exp. 2 were performed using one-tail tests.

1. Furl N, Garrido L, Dolan RJ, Driver J, Duchaine B (2011) Fusiform gyrus face selectivity relates to individual differences in facial recognition ability. *J Cogn Neurosci* 23:1723–1740.
2. Xu Y (2005) Revisiting the role of the fusiform face area in visual expertise. *Cereb Cortex* 15:1234–1242.
3. Grill-Spector KS, Sayres R, Ress D (2006) High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nat Neurosci* 9:1177–1185.
4. Pinsk MA, et al. (2009) Neural representations of faces and body parts in macaque and human cortex: a comparative FMRI study. *J Neurophysiol* 101:2581–2600.
5. Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436.
6. Pelli DG (1997) The VideoToolbox software of visual psychophysics: Transforming numbers into movies. *Spat Vis* 10:437–442.

Fig. S1. Examples of images (faces, animals, cars, planes, and scrambled matrices) used in the functional MRI (fMRI) portion of the experiment. Images of faces, animals (all four-legged animal photographs in which the animal's face was typically in a nonfrontal view and assumed less than one-eighth of the total image area), and planes were obtained from Kalanit Grill-Spector (3). The face images displayed here are similar to but not from the actual set shown to participants. Car images were taken in natural settings and in nonfrontal views. Scrambled images were created on-line during the scan by parsing a randomly selected image into 16 equally sized squares, then rearranging these squares to form a novel display of scrambled pieces. Images were square except for cars, as most of the original images had a landscape aspect ratio and we wanted to preserve the normal aspect of the cars because disruption could disproportionately impede expert judgments. Rectangular images of cars were adjusted to match the area of the other images. In the scanner room, an Avotec LCD projector showed images on a translucent plastic screen at the foot of the scanner bore. Subjects wore prism glasses to view images in their upright orientation. The stimuli were presented in the center of the screen and resized on each trial to subtend a visual angle that varied between 12.6° and 17.6° to discourage use of low-level visual information to perform the detection task.

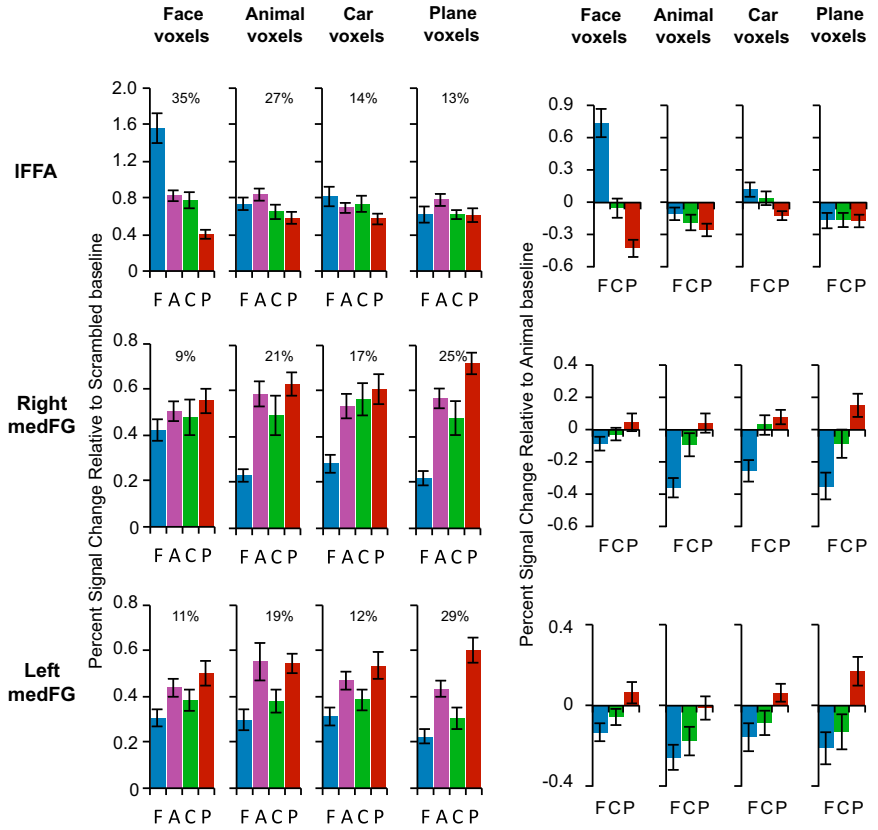

Fig. S2. Equivalent of Fig. 1 *B* and *C* for the left FFA (n = 21), right medFG (n = 19), and left medFG (n = 24). HR voxels were grouped by the category that elicited the maximal response in half of the data, and PSC for each category relative to scrambled matrices (Fig. 1*B*) or animals (Fig. 1*C*) was plotted for the other half of the data. Error bars show SEM. Percentages represent the average proportion for each kind of voxel.

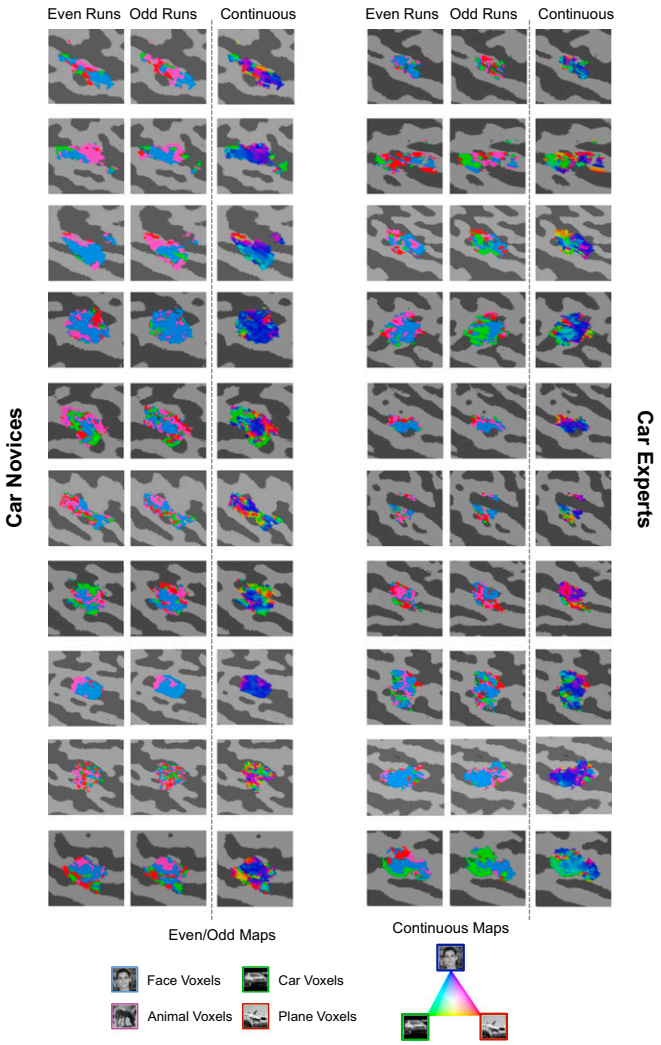

Fig. S3. Discrete and continuous activation maps. Functional results registered to the flattened cortical surface were subsampled at a resolution of 1 mm^3 over a thickness of 5 mm encompassing the white matter–gray matter boundary, and active voxels from the middle sampling depth (3 mm) are represented here. Participant maps are grouped into car novices (n = 10) and car experts (n = 10) based on a median split of car d' scores, and columns are ordered by ascending car d' values (novices: car d' = 0.52–1.65, experts: car d' = 1.74–3.17). Discrete activation maps for rFFAs show active voxels maximally selective to faces, animals, cars, or planes based on either even-run data (column 1 in each group) or odd-run data (column 2 in each group). Within the area that was face-selective with SR-fMRI, different subsets of voxels were maximally responsive to each category. A comparison across these independent samples shows reliability for certain voxels but not others. We tested the voxel-by-voxel within- and between-category correlations across even and odd runs using the Fisher transformation, an approximate variance-stabilizing transformation for r, defined by $z = 1/2ln[(1 + r)/(1 - r)]$, where ln is the natural logarithm function and r is the sample correlation coefficient. Continuous maps for rFFA for all participants are given in the third column for each group. Voxels were sorted according to their relative PSC for faces, cars, and planes using an animal baseline, as discussed in the main text.

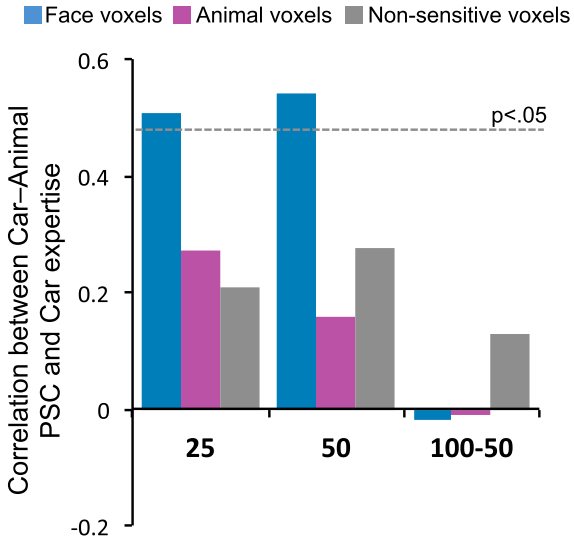

Fig. S4. Correlations in sorted voxels between PSC for cars–animals with behavioral car expertise, in circular and concentric ring ROIs: 25 mm^2, 50 mm^2, and 100–50 mm^2. The horizontal line represents the threshold for significance.

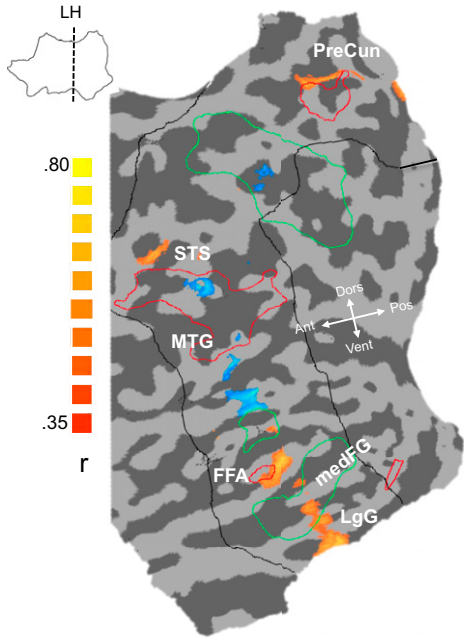

Fig. S5. Group-average map of partial correlations between PSC for cars–animals and behavioral car d' regressing out plane d', overlaid on an individual flattened left hemisphere. The black outline depicts the approximate borders of the HR field of view. Red and green outlines represent regions activated by faces > objects or objects > faces, respectively, from the group-average SR localizer data at a threshold of $P < 0.05$ with false-discovery rate correction. The most salient difference between hemispheres was an effect of car expertise in a large face-selective cluster in the right middle temporal gyrus and superior temporal sulcus (STS), but not on the left. A comparison of car expertise effects in the large ROIs from the face–object localizer found a larger correlation with car expertise on the right versus left ($r = 0.356$ vs. $r = 0.025$, $t = 1.965$, $P = 0.033$).

Table S1. Mean peak Talairach coordinates and mean size for different ROIs

Experiment	Mean Talairach coordinates for peak voxel ± SD	Average volume (mm^3) ± SD
Exp. 1		
Right FFA (*n* = 20)	37.3, −55.9, −16.7 (4.9, 9.0, 4.2)	687 (120)
Left FFA (*n* = 21)	−35.9, −53.1, −15.6 (3.0, 8.1, 3.8)	344 (87)
Right medFG (*n* = 21)	26.0, −57.1, −11.8 (4.7, 5.0, 4.1)	943 (196)
Left medFG (*n* = 24)	−27.7, −59.6, −11.4 (3.8, 7.2, 3.9)	888 (161)
Right FFA1 (*n* = 18)	34.2, −64.8, −16.1 (4.5, 7.9, 6.0)	533 (133)
Right FFA2 (*n* = 18)	35.9, −52.8, −18.5 (4.5, 6.0, 4.4)	631 (194)
Left FFA1 (*n* = 15)	−38.2, −64.9, −18.0 (3.3, 7.0, 4.7)	511 (118)
Left FFA2 (*n* = 20)	−37.5, −51.4, −17.6 (3.56, 6.1, 3.5)	433 (135)
Exp. 2		
Right FFA1 (*n* = 23)	36.0, −66.3, −16.3 (6.3, 7.4, 5.4)	685 (150)
Right FFA2 (*n* = 19)	35.2, −50.3, −14.8 (6.4, 10.0, 9.1)	677 (169)
Left FFA1 (*n* = 19)	−38.9, −64.0, −17.5 (5.2, 7.9, 5.0)	595 (204)
Left FFA2 (*n* = 18)	−39.5, −49.1, −17.4 (6.4, 7.6, 4.1)	659 (193)

These sizes were chosen on the basis of a review of the typical size of these regions in the literature (34 SR-fMRI studies) and to maximize the number of subjects for each ROI. In Exp. 1, where voxels were included in the single FFA ROI as a distance from the peak in 2D, this volume corresponds to the 100-mm^2 ROI on the right and the 50-mm^2 ROI on the left. The 25, 100, 200, and 300 ROIs reported in Fig. 3 and Fig. S4 were created by changing the threshold. The medFG ROI was much larger and allowed to cover an area as large as 200 mm^2. The FFA1/FFA2 (posterior/anterior) ROIs were defined in 3D in Exp. 1 and in 2D in Exp. 2, and allowed to reach a size no larger than ~100 mm^2 on the cortical surface. In general, ROIs are almost identical whether they are defined in 2D or 3D.

Table S2. Correlations between car expertise and the PSC for cars–animals the four groups of voxels or with category voxel selectivity (ROI d_a)

PSC and selectivity	Face	Animal	Car	Plane
PSC for cars–animals in category-selective voxels				
Right FFA	**0.486**	0.32	**0.738**	**0.545**
	0.409	0.283	**0.695**	**0.467**
Right medFG	**0.638**	**0.518**	0.365	**0.53**
	0.483	0.401	0.299	**0.594**
Left FFA	**0.596**	0.296	**0.548**	0.02
	0.437	0.219	0.411	0.072
Left medFG	0.335	0.29	0.322	0.335
	0.324	**0.499**	0.333	**0.504**
Right FFA1	**0.5**	−0.008	**0.814**	**0.802**
	0.362	−0.119	**0.669**	**0.726**
Right FFA2	**0.516**	0.27	**0.617**	0.311
	0.533	0.275	**0.6**	0.335
Left FFA1	0.181	−0.033	−0.298	−0.071
	−0.24	0.157	−0.413	0.173
Left FFA2	0.434	**0.44**	0.368	0.272
	0.426	**0.468**	0.225	0.28
Category-specific voxel selectivity, d_a				
Right FFA	−0.413	**−0.646**	**0.519**	−0.016
	−0.237	**−0.558**	0.418	−0.389
Right medFG	−0.368	**−0.592**	**0.761**	−0.039
	−0.14	**−0.486**	**0.694**	**−0.456**
Left FFA	0.28	**−0.548**	**0.625**	−0.079
	0.23	−0.386	**0.551**	−0.154
Left medFG	−0.12	**−0.423**	0.163	**−0.42**
	−0.08	**−0.478**	0.249	**−0.519**
Right FFA1	−0.126	−0.334	**0.566**	−0.003
	0.037	−0.124	**0.51**	**−0.175**
Right FFA2	**−0.557**	−0.243	0.336	−0.359
	−0.525	−0.192	0.273	**−0.46**
Left FFA1	0.071	−0.385	−0.229	−0.183
	0.131	−0.256	−0.288	−0.447
Left FFA2	−0.159	**−0.669**	**0.555**	−0.248
	−0.169	**−0.544**	0.389	−0.229

For each ROI, the zero-order (in white rows) and partial, with plane performance regressed out (in gray rows), correlations between car expertise and the PSC for cars–animals the four groups of voxels or with category voxel selectivity (ROI d_a). Note that for PSC, we break down the "nonresponsive" voxels into car- and plane-sensitive voxels. It is important to appreciate the different meaning of these two measures. For example, the PSC in face voxels represents the response to cars in voxels that respond maximally to cars in the other half of the dataset; face d_a represents the selectivity for faces relative to other categories in voxels that respond maximally to cars in the other half of the dataset. Bold entries represent significant correlations at $P < 0.05$.

Table S3. Correlations between plane expertise and the PSC for cars–animals for the four groups of voxels or with category voxel selectivity (ROI d_a)

PSC and selectivity	Face	Animal	Car	Plane
PSC for cars–animals in category-selective voxels				
Right FFA	**0.54**	**0.472**	**0.508**	**0.723**
	0.497	0.448	0.314	**0.636**
Right medFG	**0.59**	**0.528**	0.285	**0.659**
	0.394	0.366	−0.013	**0.483**
Left FFA	0.184	0.204	0.078	0.133
	0.217	0.075	0.139	0.234
Left medFG	0.002	0.112	0.193	0.332
	−0.24	−0.23	0.055	0.294
Right FFA1	0.225	0.273	0.273	**0.463**
	0.421	0.392	0.251	**0.491**
Right FFA2	0.386	−0.003	0.345	0.247
	0.199	−0.043	0.183	0.173
Left FFA1	0.096	0.025	0.018	0.212
	0.214	−0.023	0.288	0.128
Left FFA2	0.193	0.38	**0.475**	0.06
	0.166	0.23	**0.593**	0.266
Category-specific voxel selectivity, d_a				
Right FFA	−0.423	−0.397	0.347	**0.483**
	−0.256	−0.061	0.085	**0.591**
Right medFG	−0.152	−0.391	**0.434**	**0.469**
	0.178	−0.053	−0.046	**0.617**
Left FFA	0.162	**−0.463**	**0.636**	0.078
	0.004	−0.208	0.101	0.154
Left medFG	−0.089	−0.033	−0.05	0.01
	−0.018	0.258	−0.198	0.396
Right FFA1	−0.281	**−0.435**	0.286	0.249
	−0.249	−0.31	−0.038	0.293
Right FFA2	−0.204	−0.158	0.21	0.121
	0.061	−0.054	0.068	0.327
Left FFA1	−0.054	−0.312	−0.001	0.262
	−0.121	−0.103	0.177	0.47
Left FFA2	−0.013	**−0.479**	**0.446**	−0.114
	0.076	−0.141	0.128	0.03

For each ROI, the zero-order (in white rows) and partial, with car performance regressed out (in gray rows), correlations between plane expertise and the PSC for cars–animals for the four groups of voxels or with category voxel selectivity (ROI d_a). Note that for PSC, we break down the "nonresponsive" voxels into car- and plane-sensitive voxels. It is important to appreciate the different meaning of these two measures. For example, the PSC in face voxels represents the response to planes in voxels that respond maximally to cars in the other half of the dataset; face d_a represents the selectivity for faces relative to other categories in voxels that respond maximally to cars in the other half of the dataset. Bold entries represent significant correlations at $P < 0.05$.

Table S4. Group analyses on the right and left hemispheres (Fig. 4 and Fig S5, respectively) revealed regions where the neural response for cars (relative to animals) correlated with behavioral car expertise once plane performance was regressed out

Region	Mean Talairach coordinates for peak voxel	Peak correlation, *r* (*P* value)	Average volume (mm^3)
Right FFA	35, −56, −22	0.65 (0.002)	484
Right ITG	50, −65, −14	0.58 (0.008)	265
Right LgG	12, −60, −16	0.67 (0.001)	1,302
Right MTG	54, −59, 3	0.65 (0.002)	1,147
Right STS	58, −59, 24	0.70 (0.001)	1,301
Right Precuneus	5, −54, 33	0.60 (0.005)	349
Left FFA	−39, −55, −19	0.60 (0.004)	341
Left LgG	−35, −62, −4	0.76 (<0.001)	378
Left STS	−55, −65, 31	0.51 (0.02)	498
Left Precuneus	−12, −52, 31	0.56 (0.01)	313

ITG, inferior temporal gyrus; LgG, lingual gyrus; MTG, middle temporal gyrus; STS, superior temporal sulcus.