Neuron, Volume 53; Supplemental Data
Categorization Training Results in Shape- and Category-Selective Human Neural Plasticity

Supplementary Material

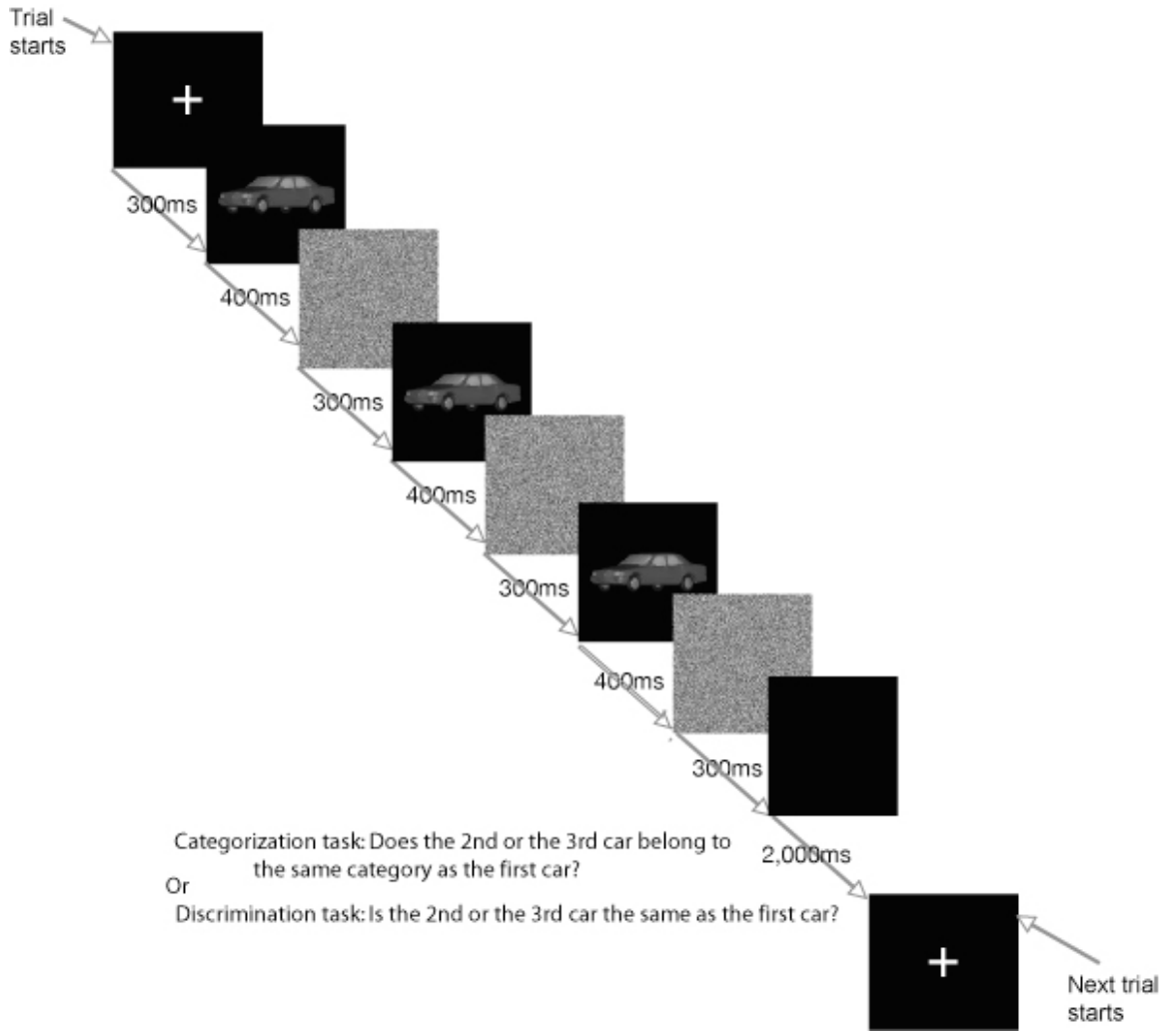
Xiong Jiang[1], Evan Bradley[1], Regina A. Rini[1], Thomas Zeffiro[2], John VanMeter[3], and Maximilian Riesenhuber[1]

[1]Department of Neuroscience, Georgetown University Medical Center, Washington, DC 20007, USA

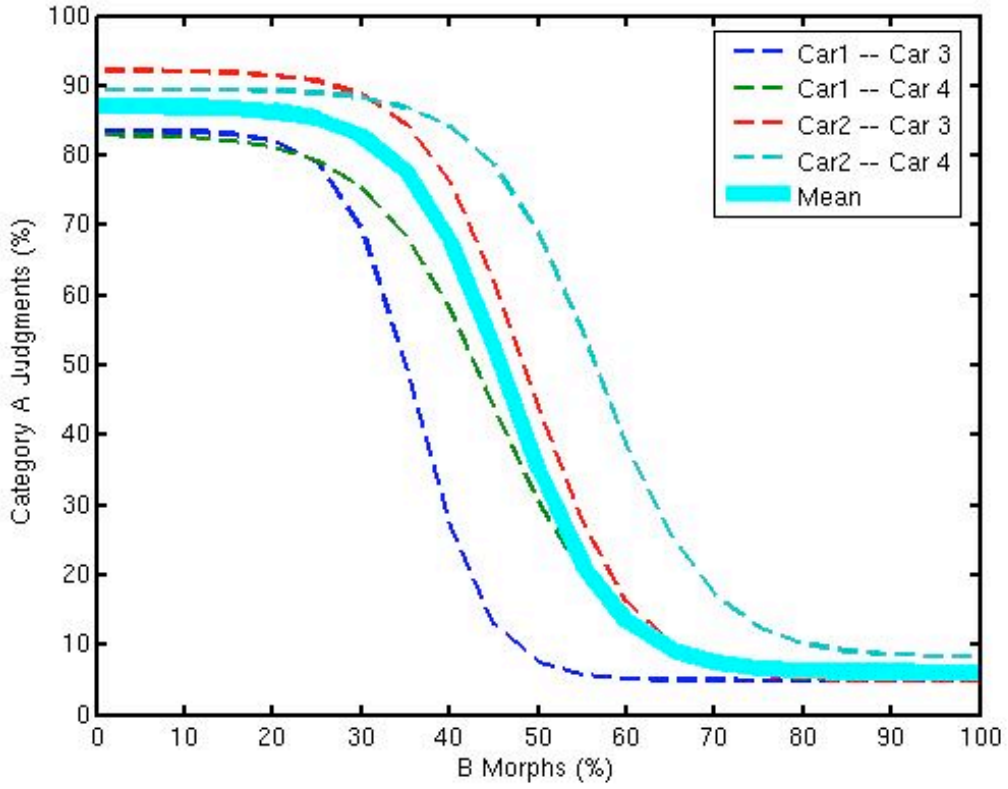
[2]Neural Systems Group, Massachusetts General Hospital, Charlestown, MA, 02129, USA

[3]Center for Functional and Molecular Imaging, Georgetown University Medical Center, Washington, DC 20007, USA

Correspondence should be addressed to M.R. (mr287@georgetown.edu).



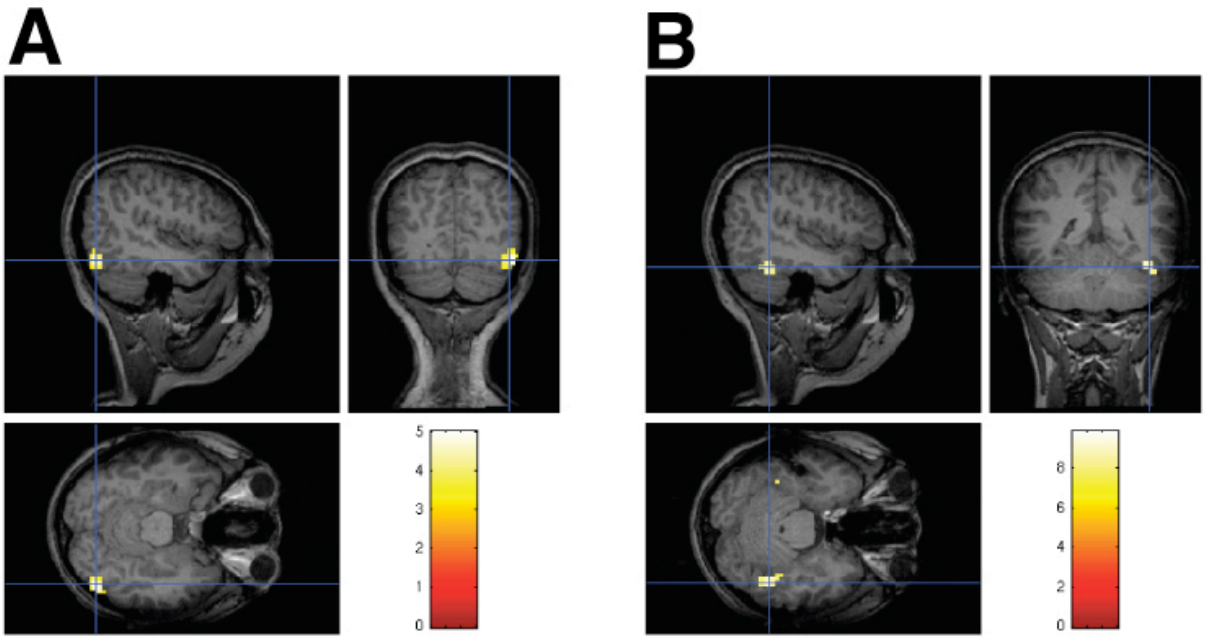
Supplementary Figure 1. Behavioral paradigm.
This is the 2AFC paradigm used in categorization training and testing, as well as in discrimination testing.



Supplementary Figure 2. Fitted behavioral categorization data.
To better capture the steep transition around the category boundary and avoid the blurring caused by averaging across subjects and morph lines, we fit the response data from each subject along each of the four morph lines to sigmoid functions,

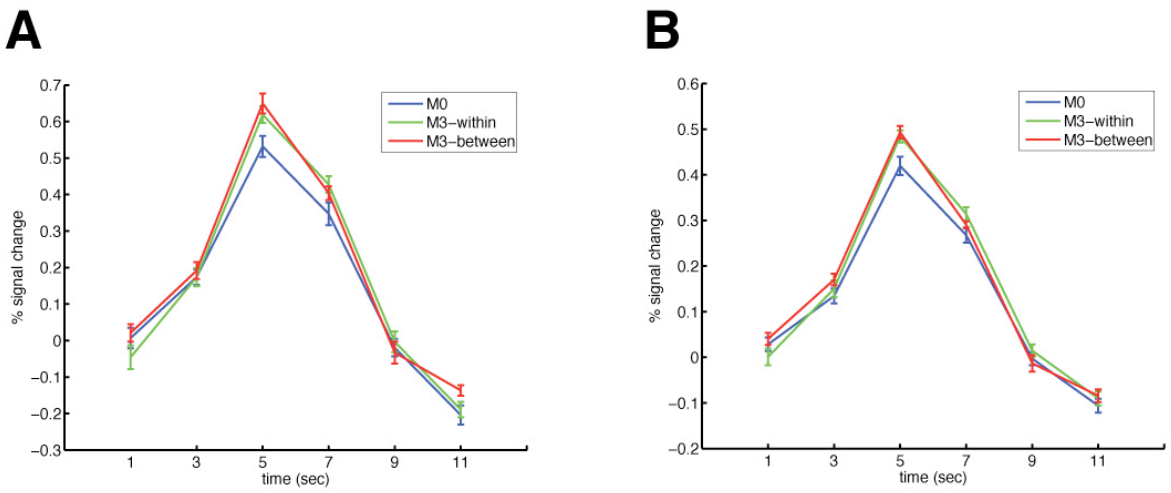
$$f(x) = \frac{a}{1 + \exp\left(\frac{-(b - x)}{t}\right)} + e$$

with *x* being the morph step position variable (ranging from 0 to 100), and *a*, *b*, *t*, and *e* parameters of the sigmoid. Among these four parameters, *b* represents the location of the category boundary, i.e., the position on the morph line where performance reaches chance level, with 50 corresponding to the ideal 50% value. The parameter *t* represents the sharpness of change around the boundary. The parameters *a* and *e* constrain the upper and lower limits of the subject's performance along the specific morph line, with ideal values 1 and 0, respectively. Optimum values for these four parameters were determined independently based on each subject's 2AFC categorization data (see Methods) along each of the four morph lines using the simplex search method implemented in the MATLAB *fminsearch* function. If *fminsearch* failed to yield a reasonable value, i.e., it returned a value for *b* that was outside the valid range (smaller than 0 or larger than 100), we manually constrained the value of *b* to lie in the range of 10-90, and redid the search, yielding fits of *a*, *t*, and *e* for every possible *b* value between 10 and 90 at a resolution of 0.5. The best fit (with minimal error) among these was picked at the end.

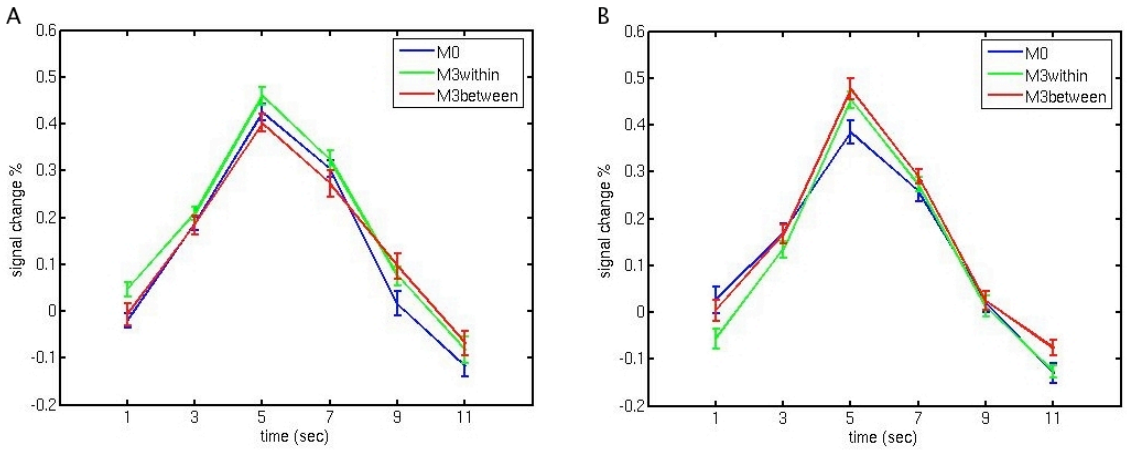


Supplementary Figure 3. The rLO and rFFA regions in one representative subject. The activations of (A) right LO (rLO) and (B) right FFA (rFFA) from one representative subject during the localizer scans. After being masked with the contrast of car or face versus baseline ($p<10^{-5}$, uncorrected), respectively, rLO was defined by the contrast of car images versus scrambled car images ($p<10^{-3}$, uncorrected for illustration purposes), and the rFFA was defined by the contrast of face versus car and scrambled car images ($p<10^{-10}$, uncorrected for illustration purposes).

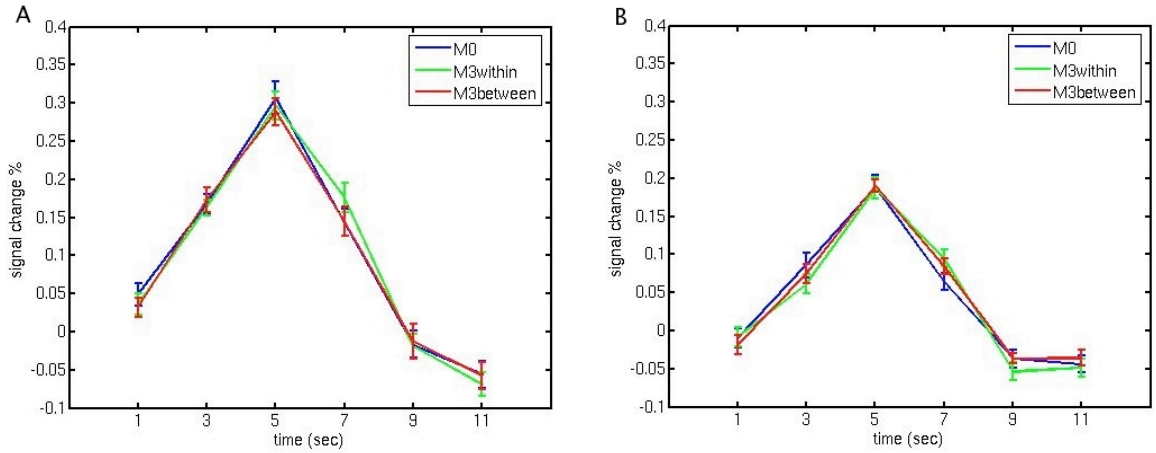
Mean sizes across subjects were 285±12mm^3 for right LO (mean MNI coordinates ± SEM, 48±0.9, -75±0.9, -11±0.8), 266±12mm^3 for left LO (mean MNI coordinates ± SEM, -43±0.9, -78±0.9, -10±1.0), 1345±81mm^3 for the right FFA (mean MNI coordinates ± SEM, 41±0.6, -50±1.4, -21±0.7), and 1046±62mm^3 for the left FFA (mean MNI coordinates ± SEM, -37±2.1, -53±1.1, -21±0.7). Here we decided to define a small LO ROI and a large FFA ROI because we hypothesized that (1) only neurons in those voxels responding strongly to cars would exhibit shape-selectivity to car images; (2) the potential effect in the FFA could either come from face-selective neurons per se (Gauthier et al., 1999), or from car-selective neurons in the overlapping pFs region (Grill-Spector et al., 2001). Because in general the responses in the right LO and FFA ROIs were not only stronger but also more sensitive to fine shape changes (see Figures S7 and S9 for left hemisphere results), we focused our data analysis for the ROIs on the right hemisphere. Furthermore, for Experiment 2, to probe the robustness of our findings we extracted two additional sets of data from rLO with a mean size of 128±5mm^3, and 545±24mm^3 (Figure S4), and similar trends were found in the two data sets.



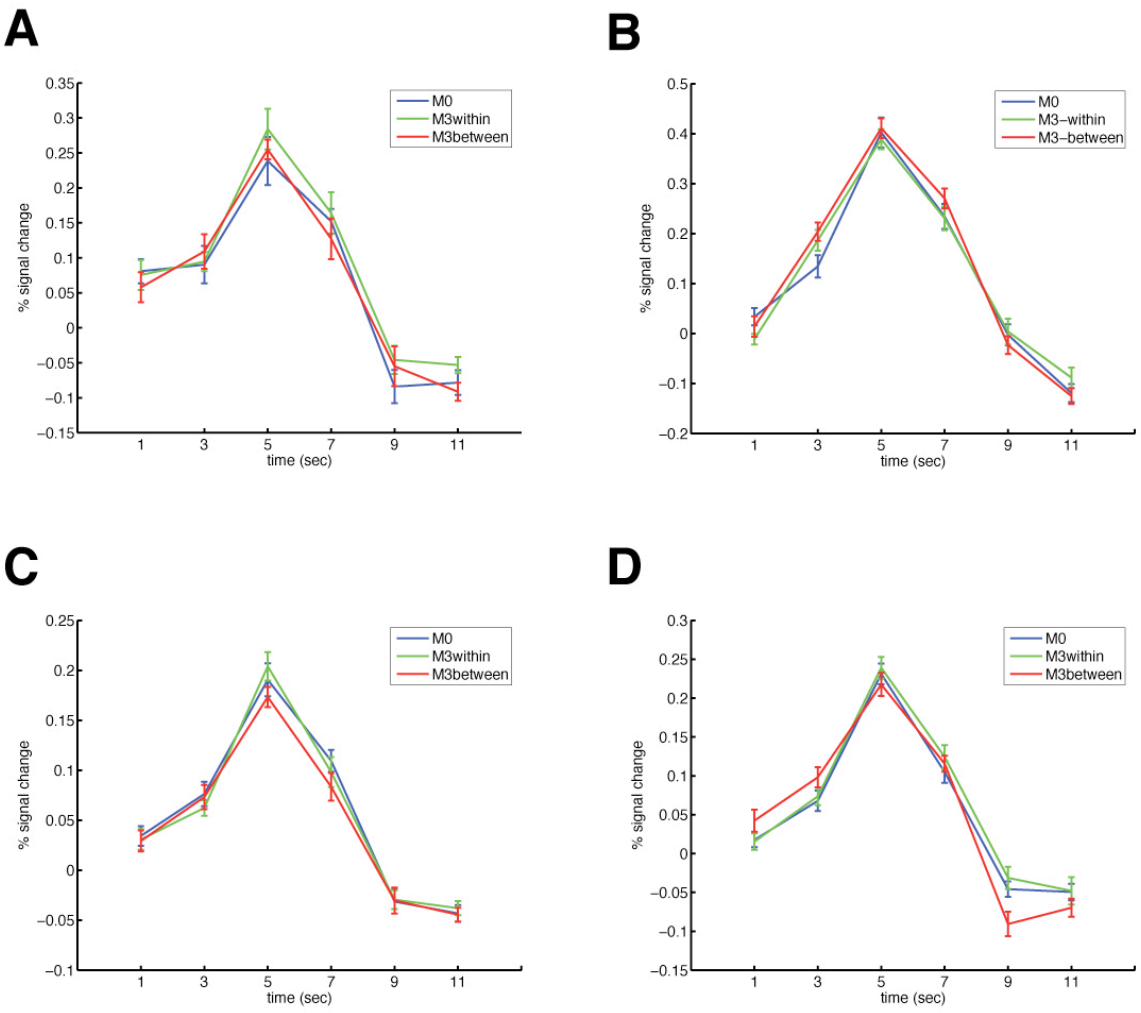
Supplementary Figure 4. fMRI responses at rLO in Experiment 2.
The mean hemodynamic responses at rLO in Experiment 2 for two different rLO ROI sizes, (A) 128±5mm^3, and (B) 545±24mm^3. For both data sets, an ANOVA with three conditions (M0, $M3_{within}$, and $M3_{between}$) as repeated measures on the peak values revealed significant differences among the three conditions, (A) $F(2, 34)=3.817$, $p<0.05$, (B) $F(2, 34)=4.510$, $p<0.05$. Error bars show within-subject SEM.



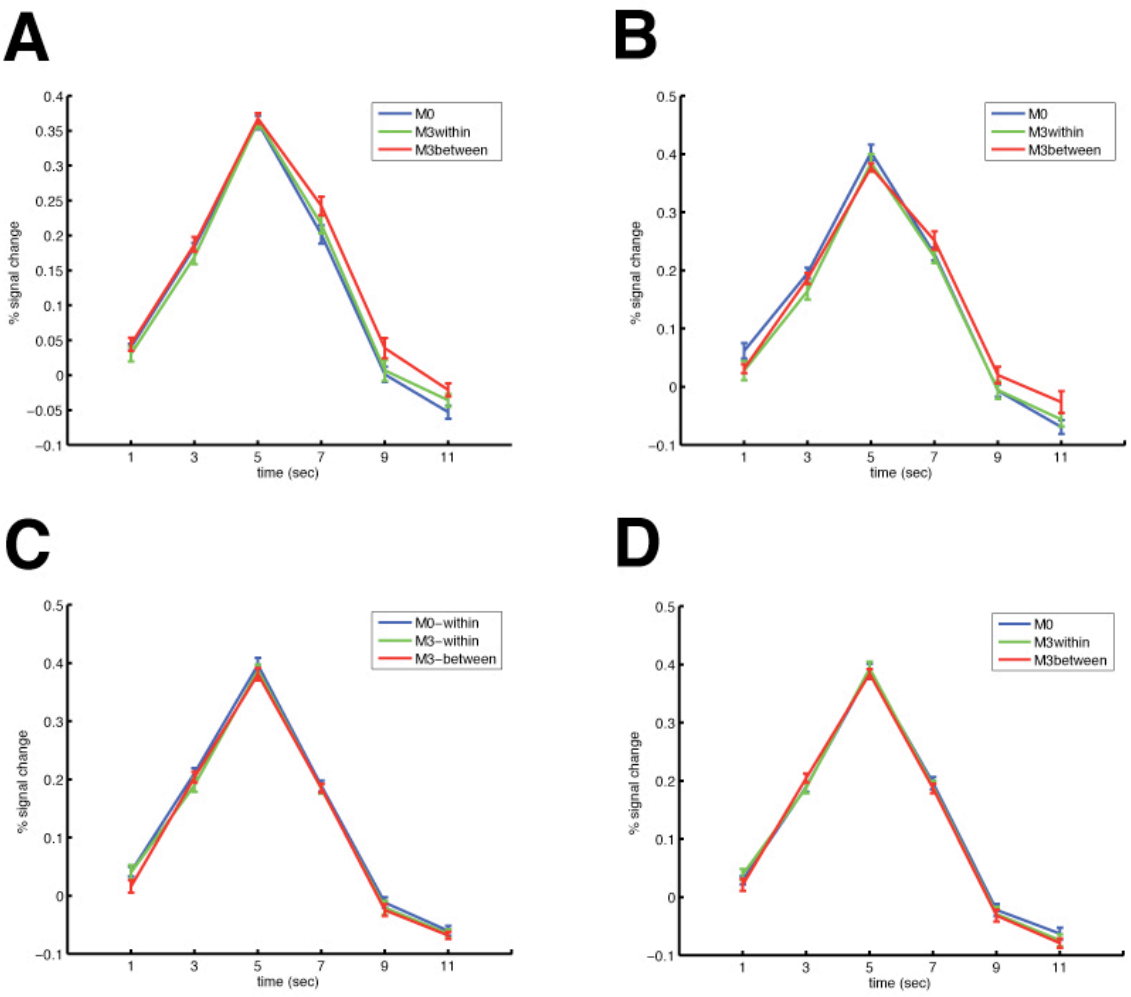
Supplementary Figure 5. fMRI responses at rLO in Experiment 1 and 2.
The mean hemodynamic responses at rLO in (A) Experiment 1 and (B) Experiment 2 from the subjects participating in both experiments (n=15). A repeated measures ANOVA conducted with following factors: training (pre- vs. post-training) and conditions (M0, $M3_{within}$, and $M3_{between}$) found a significant interaction between training and morph conditions, $F(2,28)=4.518$, $p < 0.05$, but no significant training effect ($p>0.5$), and no significant difference among the three morph conditions ($p>0.1$). Three additional ANOVA tests with factors of training (pre- vs. post-training) and morph conditions (M0 vs. $M3_{within}$, M0 vs. $M3_{between}$, and ($M3_{within}$ – M0) vs. ($M3_{between}$ – M0)) revealed a significant interaction between training and morph conditions for the comparison of M0 vs. $M3_{between}$ but not for the M0 vs. $M3_{within}$, comparison. However, this difference is likely due to the slightly (ns) higher activation for the $M3_{within}$ condition (compared to both $M3_{between}$ and M0 conditions) pre-training and not due to differential training effects, as there was no significant interaction in the ($M3_{within}$ – M0) vs. ($M3_{between}$– M0) ANOVAs nor any difference between the M3 conditions post-training (see main text). Error bars show within-subject SEM.



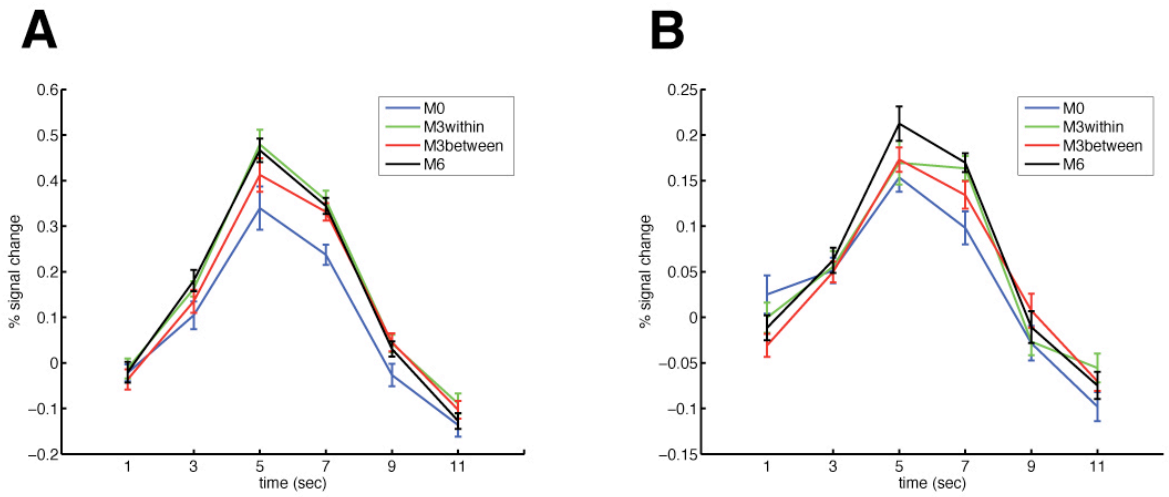
Supplementary Figure 6. fMRI responses at rFFA in Experiment 1 and 2.
The mean hemodynamic responses at rFFA in (A) Experiment 1 and (B) Experiment 2 from the same subgroup of subjects (n=15). A repeated measures ANOVA was conducted with the following factors: session (pre- vs. post-training) and conditions (M0, $M3_{within}$, and $M3_{between}$). A significant effect of session was found, $F(1,14) = 11.789$, $p<0.005$, but no significant difference was seen among the three conditions, $p>0.5$, and no significant interaction between session and morph conditions was seen, $p>0.5$. The decreased fMRI responses to cars could suggest that neurons in the FFA coding for non-car stimuli were suppressed when viewing cars after training. Error bars show within-subject SEM.



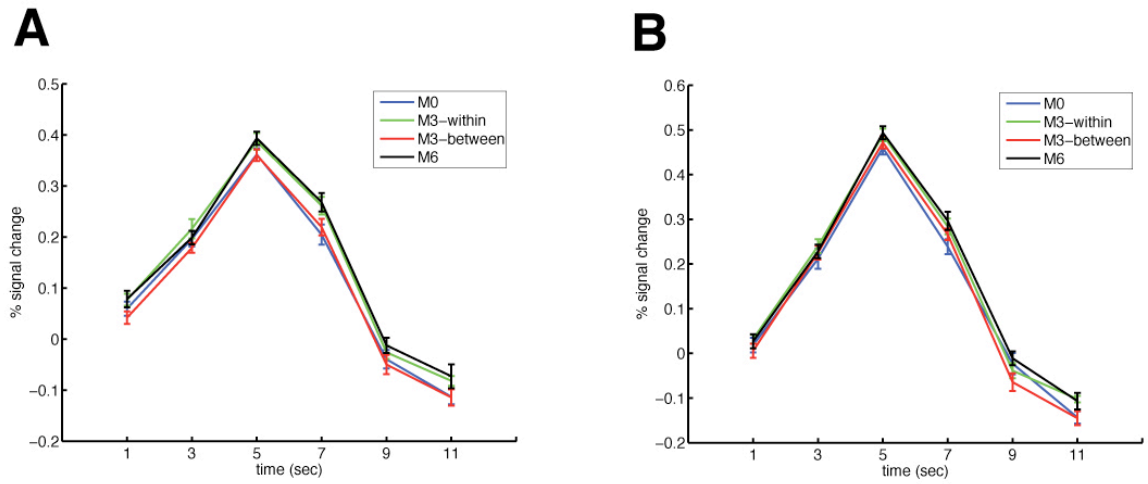
Supplementary Figure 7. fMRI responses at left LO and FFA in Experiments 1 and 2. The mean hemodynamic responses at left LO (A) before and (B) after categorization training, and at left FFA (C) before and (D) after categorization training. ANOVA with three conditions (M0, $M3_{within}$, and $M3_{between}$) as repeated measures revealed no difference at the peaks for any of the four data sets. Error bars show within-subject SEM.



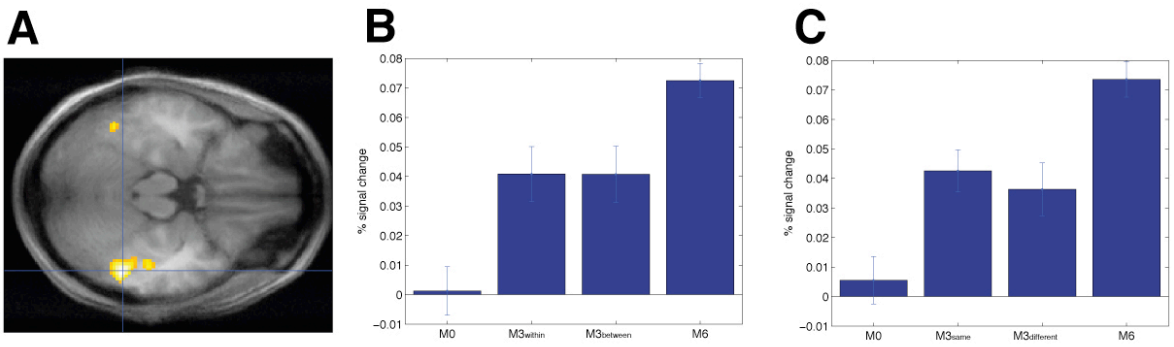
Supplementary Figure 8. fMRI responses in early visual areas in Experiments 1 and 2. The mean hemodynamic responses at the V1/V2 regions (mean size in all three experiments, right: 8229±628mm^3, left: 7778±562mm^3) from Experiment 1 (A) left, (B) right, and Experiment 2 (C) left, (D) right. No significant difference of the peaks among the three conditions (M0, $M3_{within}$, and $M3_{between}$) was found for any of the four data sets. Error bars show within-subject SEM.



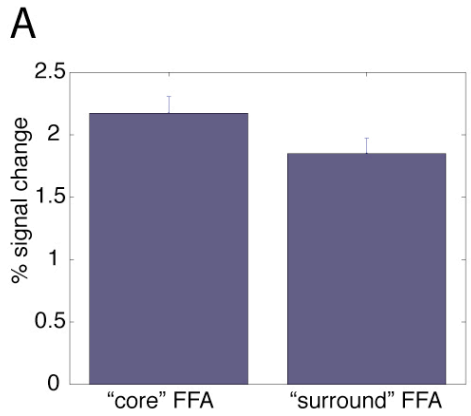
Supplementary Figure 9. fMRI responses at left LO and FFA in Experiment 3.
The mean hemodynamic responses at (A) left LO and (B) left FFA from Experiment 3. For peak values (mean of 3rd and 4th TR), ANOVA with four conditions (M0, $M3_{within}$, $M3_{between}$ and M6) as repeated measures found a significant difference at left LO, $F(3, 45)=3.987$, $p<0.05$, and at left FFA, $F(3, 39)=3.236$, $p<0.05$. For left LO, post-hoc t-test revealed significant differences between M0 and $M3_{within}$ ($p<0.05$), between M0 and M6 ($p=0.01$), and between M0 and the mean of $M3_{within}$ and $M3_{betwen}$ ($p<0.05$), but not between any other comparisons. For left FFA, post-hoc t-tests revealed significant differences between M0 and M6 ($p<0.01$), and between M0 and the mean of $M3_{within}$ and $M3_{between}$ ($p<0.05$), but not between any other comparisons. Error bars show within-subject SEM.

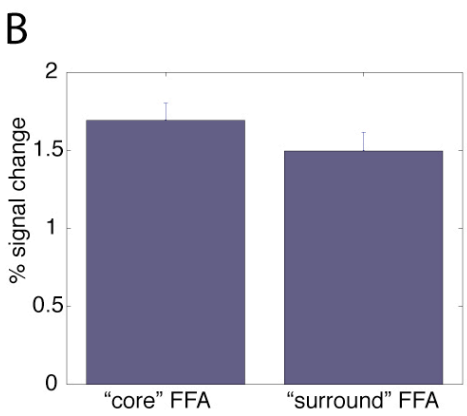


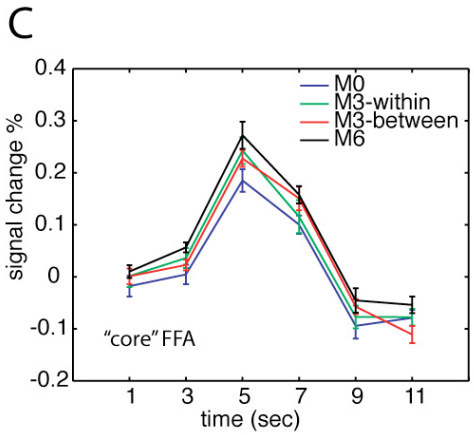
Supplementary Figure 10. fMRI responses in early visual areas in Experiment 3. The mean hemodynamic responses at (A) left and (B) right V1/V2 from Experiment 3. No significant difference at peak values among the four conditions (M0, $M3_{within}$, $M3_{between}$, and M6) was found for either data set. Error bars show within-subject SEM.

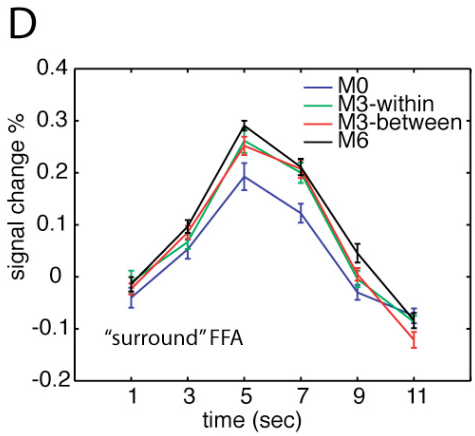


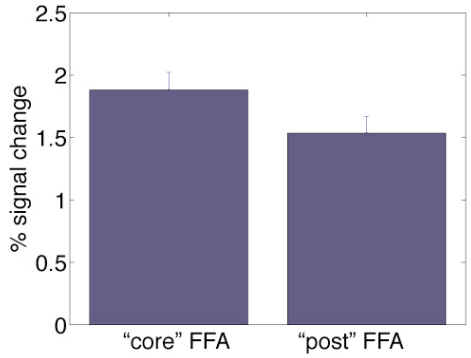
Supplementary Figure 11. fMRI responses at right pFs.
(A) The right pFs ROI was defined by the comparison of M6 versus M0 at group level ($p<0.001$, uncorrected, size: 1608mm^3, MNI coordinates: 50, -48, -20). (B), and (C) show activation in this ROI from 15 subjects (*the data from one subject were excluded from this analysis because this subject's right FFA (and thus pFs as well) was rather posterior, and the inclusion of this subject's data significantly increased variations, though did not change the main conclusion and overall pattern).* (B) shows the mean signal change for the four predefined conditions (M0, $M3_{within}$, $M3_{between}$, and M6) in this ROI. ANOVA with four conditions (M0, $M3_{within}$, $M3_{between}$, and M6) as repeated measures revealed a significant difference, $F(3, 42) = 9.826$, $p<0.001$. Post-hoc t-tests revealed significant differences between all pairs, except between $M3_{within}$ and $M3_{between}$ ($p>0.9$). (C) Because the group analysis did not find a compatible ROI at the LO region that showed such a stronger activation to M6 than to M0 (Table 1), it suggested that the adaptation effect might be stronger in pFs than in LO, in line with previous reports (e.g., Grill-Spector, et al., *Neuron,* 1999), However, such a strong difference between M6 and M0 in the pFs region also raised the question whether the pFs might encode categorical information (see Rotshtein et al. *Nature Neuroscience*, 2005; Jiang et al., *Neuron*, 2006). To address this question, we conducted an ANOVA with four post-defined conditions (M0, $M3_{same}$, $M3_{different}$, and M6, see Methods) as repeated measures, and found a significant difference, $F(3, 42) = 10.081$, $p<0.001$. Similarly as in (B), post-hoc t-tests revealed significant differences between all pairs, except for a marginal effect between M0 and $M3_{different}$ ($p=0.06$), and not between $M3_{same}$ and $M3_{different}$ ($p>0.6$). Contrasting these results to the data from right lateral PFC (see Figures S14 and Figure 6 in main text), strongly suggests that the neurons in the pFs regions just carry shape information, but no explicit category information, in line with our recent report (Jiang et al., *Neuron*, 2006). Error bars show within-subject SEM.

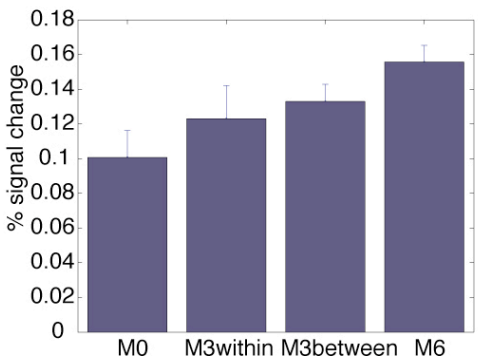

A
% signal change
2.5
2
1.5
1
0.5
0
"core" FFA
"surround" FFA

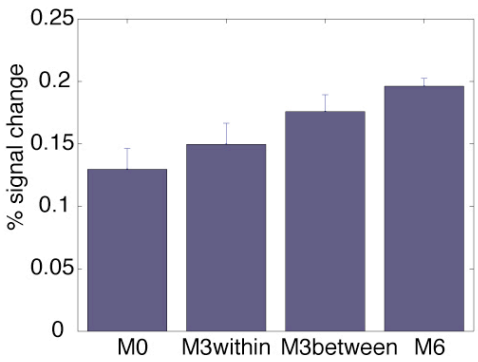

B
% signal change
2
1.5
1
0.5
0
"core" FFA
"surround" FFA


C
signal change %
0.4
0.3
0.2
0.1
0
-0.1
-0.2
M0
M3-within
M3-between
M6
"core" FFA
1
3
5
7
9
11
time (sec)


D
signal change %
0.4
0.3
0.2
0.1
0
-0.1
-0.2
M0
M3-within
M3-between
M6
"surround" FFA
1
3
5
7
9
11
time (sec)


E
% signal change
2.5
2
1.5
1
0.5
0
"core" FFA
"post" FFA


F
% signal change
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
M0
M3within
M3between
M6


G
% signal change
0.25
0.2
0.15
0.1
0.05
0
M0
M3within
M3between
M6

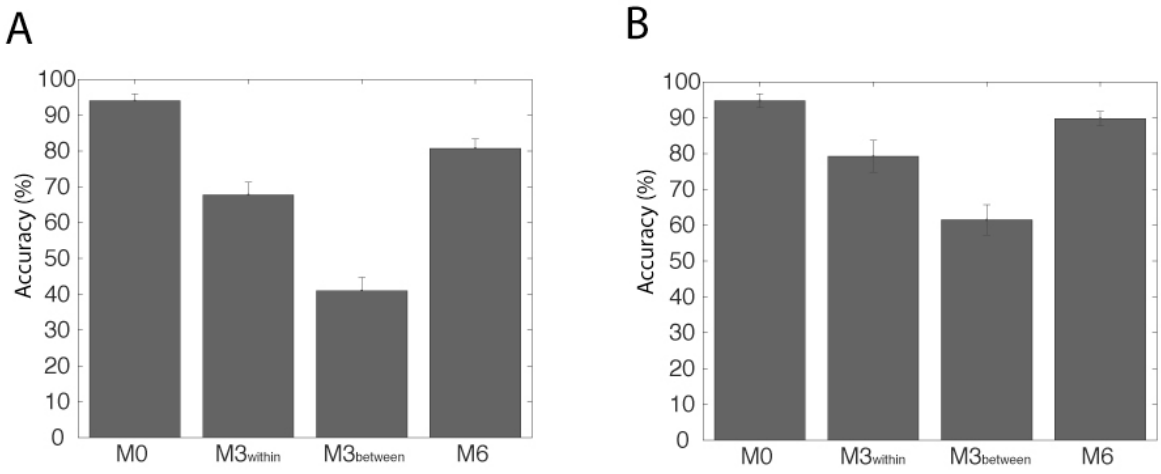
Supplementary Figure 12 (preceding page). fMRI responses in the "core" and "surround" FFA ROIs.

(A) shows the mean signal change in the "core" face and "surround" face ROIs (see Methods) to faces from the localizer scans when the FFA is defined by the contrast of face vs. cars. Paired t-tests revealed that the "core" face ROI responded more strongly to faces than the "surround" face ROI ($p<0.00005$). Mean size of the face ROIs was $679\pm101mm^3$ for the "core FFA" ROI and $631\pm90mm^3$ for the "surround FFA" ROI.

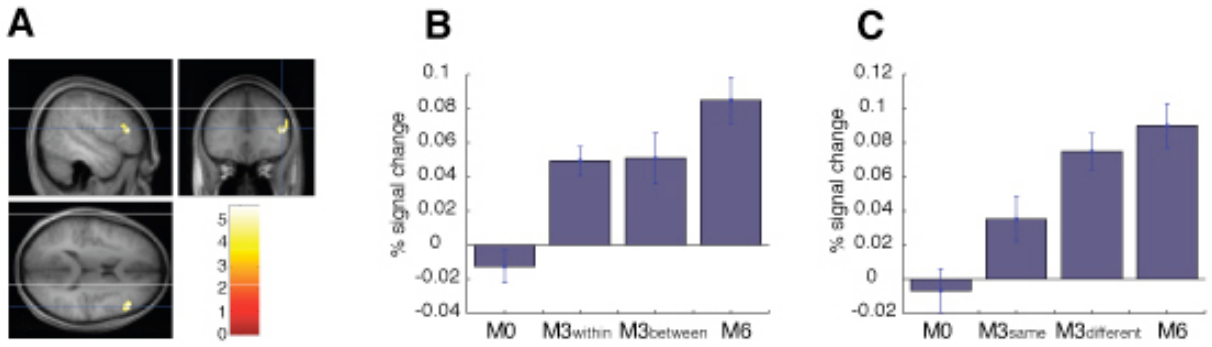
(B) To directly address the hypothesis that the FFA might mediate expertise learning, with individual neurons in the FFA coming to respond to *both* faces and cars (Gauthier, *TICS,* 2000) we redefined the FFA by the contrast of face vs. scrambled cars. (B) shows the mean signal change to faces from the localizer scans in the "core" face and "surround" FFA ROIs derived from the redefined FFA in an analogous fashion as the corresponding ROIs in (A). Paired t-tests revealed that the "core" face ROI responded marginally more strongly to faces than the "surround" face ROI ($p<0.05$, one-tailed). (C) We then extracted the mean hemodynamic responses to cars in Experiment 3 in the redefined "core FFA" and "surround FFA" ROIs from (B). ANOVAs with the factors conditions (M0, M3within, M3between, and M6), and ROI ("core" and "surround"), did not find significant effects of conditions or ROI, nor an interaction. However, running ANOVAs on the data sets from the two ROIs separately found a marginal effect in the "surround" ($F(3,45)=3.109$, $p=0.06$), but not in the "core" FFA ($p>0.5$).

In a final analysis, we defined ROIs relative to fixation to gain the broadest possible FFA definition. Specifically, we redefined the "core" FFA as a sphere (radius = 4mm) centered at the peak of the face versus fixation contrast from the localizer scans, and then defined another ROI, termed "post" FFA, as another sphere with the same size but centered 6mm more posterior and 2mm more lateral than the "core" FFA, with the specific displacement chosen based on the spatial distance between the locations of peak activation in the FFA (to faces) and pFs (to cars) from the group data of the localizer scans, in agreement with the literature on the locations of FFA and pFs (e.g., Grill-Spector et al., *Vision Research*, 2001). (E) shows the mean signal change in the "core" face and "post" face ROIs to faces from the localizer scans. Paired t-tests revealed that the "core" face ROI responded more strongly to faces than the "post" face ROI ($p<0.0005$). The signal change to cars in Experiment 3 under four different conditions (M0, M3within, M3between, and M6) is shown in (F) for the "core" FFA, and (G) for the "post" FFA (note that the plots shows percent signal change obtained from the standard canonical hemodynamic response function since the definition of the "post" FFA was based on group data, similar to the rLPFC ROI in Figure 6 in the paper, rather than individually identified ROI such as the FFA or LO ROIs in Figures 3 and 5 in the paper). ANOVAs with the factors conditions (M0, M3within, M3between, and M6), and ROI ("core" and "post" FFA), did not find significant effects of conditions or ROI, nor an interaction between the two. However, running ANOVAs on the data sets from the two ROIs separately again found a significant effect in the "post" FFA, which presumably overlapped with pFs, ($F(3,45)=3.480$, $p<0.05$), but not in the "core" FFA ($p>0.14$).

These analyses thus further support that the difference between M6 and M0 found in the fusiform gyrus in the voxel-wise analysis was not specific to the FFA. Error bars show within-subject SEM.

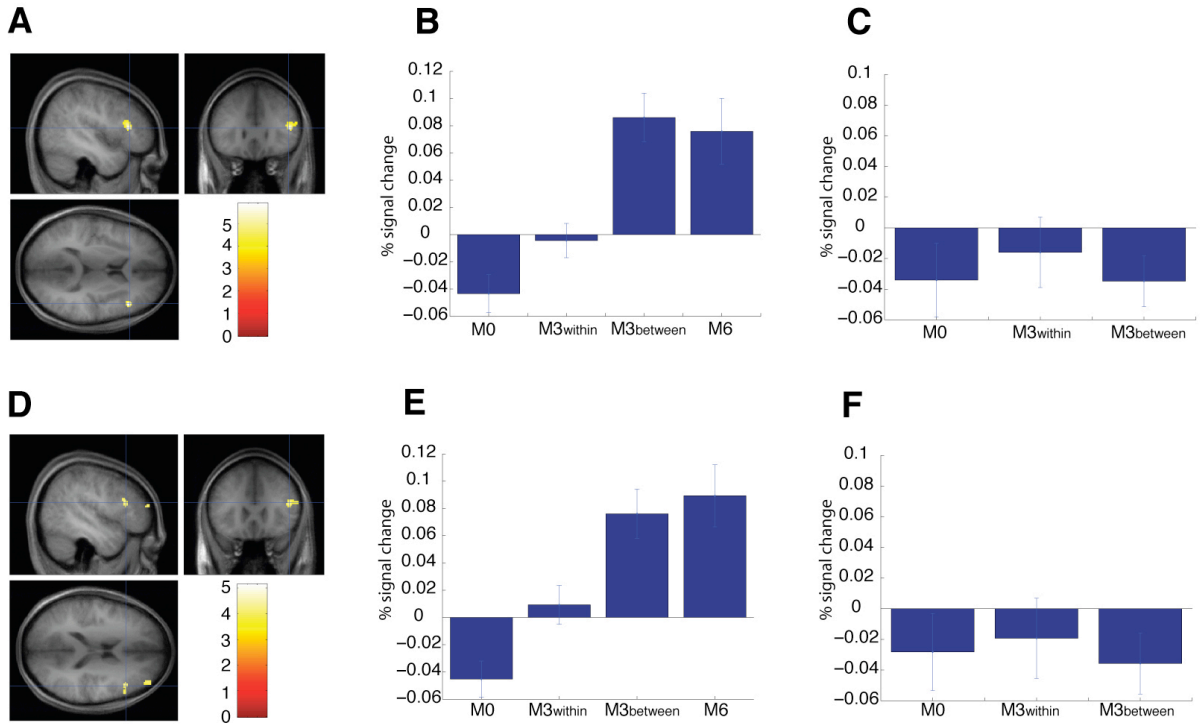


Supplementary Figure 13. Categorization performance in Experiment 3.
The mean accuracy of categorization judgments in Experiment 3 while subjects were inside the scanner. (A) shows performance averaged over all four morph lines, (B) shows performance on the "best" morph line, *i.e.,* the one on which performance was highest out of the four. Error bars show SEM.



Supplementary Figure 14. Activation in the rLPFC ROI affected by subjective categorization judgments.

(A) The rLPFC ROI defined by the comparison of M6 and $M3_{between}$ versus $M3_{within}$ and M0 across all four morph lines ($p<0.0005$, uncorrected, size: $1200mm^3$). (B) The mean signal change at this ROI in Experiment 3 for the four pre-defined conditions, M0, $M3_{within}$, $M3_{between}$, and M6. (C) To probe response-related changes, based on participants' same/different responses, M3 conditions were regrouped as $M3_{same}$ and $M3_{different}$, resulting in design matrices set up as M0, $M3_{same}$, $M3_{different}$, and M6 instead of M0, $M3_{within}$, $M3_{between}$, and M6. Repeated-measures ANOVA revealed a significant difference among the four conditions for both data sets ($p<0.0001$ each). However, post-hoc paired t-test revealed that (B) while the difference between $M3_{within}$ and $M3_{between}$ was not significant ($p>0.5$), grouping trials according to each subject's perceived category membership (C) showed a significant difference between $M3_{same}$ and $M3_{different}$ ($p<0.05$). The results thus suggest that activation in rLPFC was related to subjects' individual category membership judgments. Error bars show within-subject SEM.



Supplementary Figure 15. Activation in the rLPFC ROI affected by tasks.
(A) The rLPFC ROI defined by the comparison of M6 and $M3_{between}$ versus $M3_{within}$ and M0 of the morph line on which subjects had the best performance (see Table 2, $p<0.001$, uncorrected, size: $1088mm^3$), and mean signal change at this ROI for the conditions of M0, $M3_{within}$, $M3_{between}$, and M6 on the morph line with the best behavioral performance in the scanner in Experiment 3 (B), and for the conditions of M0, $M3_{within}$, and $M3_{between}$ in Experiment 2 (**C**). ANOVA with three conditions (M0, $M3_{within}$, and $M3_{betweeen}$) as repeated measures found a significant difference for the data set of Experiment 3 (B), $p<0.00001$, but not for the data set of Experiment 2 (C), $p>0.5$. (D) The rLPFC ROI defined by the comparison of M6 versus M0 of the same morph line as (A) (see Table 2, $p<0.001$, uncorrected, size: $512mm^3$), and mean signal change at this ROI for the conditions of M0, $M3_{within}$, $M3_{between}$, and M6 on the morph line with the best behavioral performance in the scanner in Experiment 3 (E), and for the conditions of M0, $M3_{within}$, and $M3_{between}$ in Experiment 2 (F). ANOVA with three conditions (M0, $M3_{within}$, and $M3_{betweeen}$) as repeated measures found a significant difference for the data set of Experiment 3 (E), $p<0.0001$, but not for the data set of Experiment 2 (F), $p>0.5$. Error bars show within-subject SEM.

The consistent lack of category related activations in LPFC during the position task across all comparisons suggests that it is unlikely that subjects were doing a covert categorization task while judging the position of stimuli during Experiment 2. While having subjects execute a more difficult "orthogonal" task (i.e., one unrelated to the categorization task of interest), rather than the simple position task we chose, would appear to offer a way to avoid the possibility of covert categorization, as it would come at the risk of learning effects associated with the orthogonal task potentially confounding the main learning effect. Moreover, previous studies have shown that even quite complex categorization tasks (such as "animal/no animal" (Li, van Rullen, Koch, Perona, *PNAS*, 2002) or face gender (Reddy, Wilken, Koch, *J Vis,* 2004) can be performed

concurrently with difficult, attention-demanding tasks, suggesting that covert categorization can also occur for difficult orthogonal tasks. In our study, we therefore decided to go with a simple task and attempted to minimize the possibility for covert categorization by always asking subjects to perform the position task in the scanner up to the final scan during which they performed the categorization task, and the aforementioned absence of category-selective LPFC activation suggests that this approach was successful (within the resolution of the present study). Further, given that there were at least two days between the conclusion of category training and the post-training position task scan (see Supplementary Table 4), it appears unlikely that the absence of category selectivity in the post-training position task scans but its presence in the following categorization scans could be due to memory consolidation effects.

It is further worth pointing out that the category-selective activation in the LPFC ROI in Experiment 3 cannot be an artifact of the increased task difficulty of the categorization task relative to the position task in Experiment 2, as the M6 condition, which was the easiest condition in the categorization task (apart from the trivial M0 condition, see Figure S13), evoked the highest response in the ROI while responses in the M0 and $M3_{within}$ conditions stayed low.

Supplementary Table 1. Brain regions that were strongly activated by the stimuli.

Region	mm^3	T	MNI Corrrdinates X	Y	Z
Stimuli > baseline					
Left Visual Cortex	26912	10.53	-44	-88	-8
		9.26	-22	-102	-2
		7.13	-26	-92	-14
Right Visual Cortex	33024	9.06	26	-105	4
		8.70	32	-100	8
		8.55	20	-100	-6
Left Inf/Sup Parietal	12280	8.25	-30	-60	50
		7.80	-38	-42	44
		7.00	-24	-68	48
R/L Sup Mot	7504	7.45	4	6	52
		7.12	-6	2	56
		6.94	-2	12	48
R Cerebelum	544	5.12	18	-64	-46
R Sup Parietal	1232	5.12	28	-66	54
		4.35	26	-52	48
L Brainstem	1584	4.93	-16	-18	12
		4.48	-14	-16	2
L Thalamus	216	4.85	-10	-26	-12
R Thalamus	728	4.78	16	-18	8
L Cerebelum	528	4.76	-16	-54	-20
L Precentral	656	4.65	-48	6	32
		3.87	-50	0	42
R Mid/Sup Frontal	408	4.65	32	-2	58
L Mid Frontal	240	4.60	-50	32	32
R Inf Parietal/Supramarginal	328	4.59	44	-36	46
L Sup/Mid Frontal	1008	4.57	-22	-8	44
		4.07	-26	-4	54
L Insula	568	4.49	-32	20	2
L Cerebum	280	4.46	-6	-28	26
		3.94	-8	-20	28
L Cerebelum	368	4.45	-16	-64	-50
R Cerebelum	480	4.43	36	-18	54
L Cerebelum	472	4.43	-8	-74	-20
L Rolandic	264	4.40	-46	-6	10
R Postcentral	208	4.34	62	-18	50
R Sup/Mid Occipital	760	4.33	28	-74	32

Supplementary Table 2. Brain regions that showed stronger activation to pairs of cars belonging to different categories than to those belonging to the same category, even when RT-related activation components were removed.

Since reaction times (RT) in the M0 condition were significantly shorter than those in the other three conditions in Experiment 3 ($p<0.001$, with no significant difference among the other three conditions), to rule out the possibility that the category-selective activation patterns were caused by differential RT per se, we conducted an additional analysis to explicitly model the RT-related activation (see Methods). We found that both the contrast of M6 versus M0 and the contrast of M6 and $M3_{between}$ versus $M3_{within}$ and M0 yielded very similar results (Table S2), albeit with slightly less significance, as RT might be correlated with the category-related activity itself. In contrast, RT-related activations were found in parietal cortex, motor cortex, prefrontal cortex, insular cortex, and the thalamus (Table S3), regions related to the decision evaluation of categorization, in line with a recent report (Grinband et al., 2006).

			MNI Corrrdinates		
Region	mm^3	T	X	Y	Z
M6 > M0					
R Mid Cingulum	496	6.49	16	-24	34
R Inf/Mid Frontal Gyrus	2144	5.63	44	32	16
		4.45	54	36	16
		4.32	36	36	8
R Sup/Mid Occipital	2216	5.4	34	-78	34
			28	-72	46
L Inf/Sup Parietal	4600	5.34	-24	-52	38
		5.06	-30	-52	46
		4.87	-38	-40	38
R Inf Temporal	720	5.32	50	-48	-20
L Precentral	272	5.19	-40	2	34
L Brainstem	216	5.08	-8	-22	-42
R Inf Parietal/Angular	2432	5.04	34	-46	34
		4.74	44	-42	46
		4.7	30	-54	50
R Inf Temporal	256	4.82	46	-30	-22
L Cerebrum	200	4.65	-44	-8	-18
M6 & M3between > M3within & M0					
R Inf Frontal Gyrus	520	4.41	50	32	16
		4.26	54	30	24

Supplementary Table 3. Brain regions that showed activation correlated with RT.

Region	mm^3	T	MNI Corrrdinates X	Y	Z
RT related activations*					
L Inf/Sup Parietal	5464	10.67	-36	-48	38
		7.97	-32	-62	52
R Cerebelum	2856	7.64	-22	-68	48
		9.59	28	-56	-36
		9.1	32	-62	-32
L Sup Motor Area	6872	8.86	44	-66	-26
		9.53	-6	20	42
		8.79	-2	14	52
L Insula	1728	8.62	10	18	38
		9.49	-32	18	6
R Thalamus	2104	7.42	-38	22	-4
		8.98	12	-20	10
		7.29	8	-22	-4
R Inf Parietal	1200	6.44	18	-2	18
		8.29	40	-50	46
		6.96	34	-56	46
R Insula	1880	6.8	42	-42	46
		8.25	30	24	-2
L Supremarginal	1272	7.96	38	20	8
		7.95	-56	-24	36
		7.44	-60	-22	44
		6.28	-62	-32	50
L Thalamus	512	3.86	-10	-24	10
L Cerebelum/L Inf Occipital	1368	7.83	-46	-68	-22
		7.39	-36	-62	-32
		6.79	-46	-76	-16
L Visual Cortex	280	7.55	-24	-100	-14
R Visual Cortex	232	7.26	38	-94	-8
R Inf Frontal	160	6.83	46	10	30

** p<0.00001 uncorrecte in this table*

Supplementary Table 4. Schedule of experiments and training sessions.

Time (hours)	Sessions
1	fMRI Experiment 1: position task, pre-training
1-2	Discrimination testing
4-10	Categorization training
1	Categorization testing
1	Discrimination testing
1	fMRI Experiment 2: position task, post-training
1	fMRI Experiment 3: categorization task, post-training

Supplementary Study

To show that the observed behavioral and fMRI effects were not due to test/retest effects, we performed a series of behavioral and fMRI control studies.

Methods

Participants

Sixteen (7 female, aged 18-32) normal right-handed members of the Georgetown University community participated in this experiment. Experimental procedures were approved by Georgetown University's Institutional Review Board, and written informed consent was obtained from all subjects prior to the experiment. Six of the sixteen subjects only participated in the first and second discrimination testing that did not involve training and fMRI scanning, and the other ten subjects participated in all behavioral testing, fMRI experiments, and category training, Among those ten subjects, the data from three participants were excluded because of dropping out of the study (n=1), failing to reach the pre-set learning criterion (n=1), or failing to follow instructions while inside the scanner during post-training scans (no response to more than 30% of trials, n=1).

Experimental Design

This control study was based on the same design as the study described in the main paper, though with several notable changes:

(1) To probe whether the increase in discrimination task performance was due to categorization training or was a test/retest effect (i.e., merely due to the repeated testing on the discrimination task), all sixteen subjects were tested twice with the discrimination task before training.

(2) To probe whether the increase in sensitivity to shape change in LO was due to categorization training or a test/retest effect, the subjects who also participated in the fMRI experiments were scanned twice before training while doing the position task inside the scanner (similar to a double repeat of Experiment 1 described in the main paper, with differences as described next).

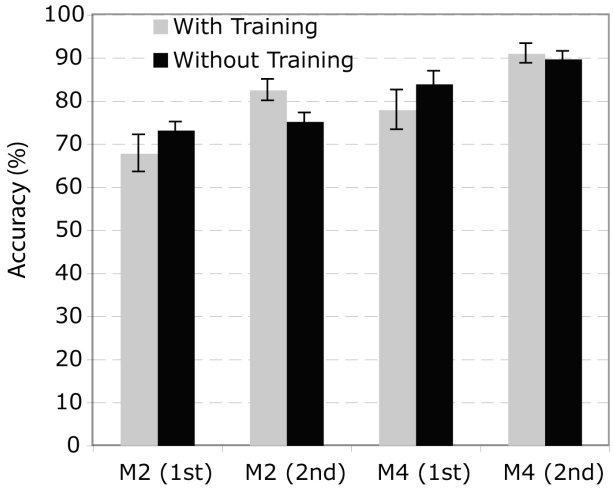
(3) The position task that subjects performed inside the scanner was changed from the previous left/right judgment to a same/different judgment (i.e., whether the first and second cars were presented at the same or different positions), to correspond to the same/different component of the categorization task (where subjects had to judge whether the first and second car belonged to the same or different categories). This change allowed us to probe whether the frontal activations previously observed during the categorization task were part of a general network of brain areas engaged in same/different tasks, or were specific to the categorization task. Accordingly, the conditions were defined as $M0_{\text{same position}}$, $M0_{\text{different position}}$, $M3_{\text{same position}}$, and $M3_{\text{different position}}$ (the $M3_{\text{within}}$ and $M3_{\text{between}}$ stimulus pairs were grouped together with equal number of repeats to boost sensitivity for the M0 vs. M3 difference of interest). The four conditions and the null trials were also counter-balanced using M-sequences, resulting in a design similar to fMRI Experiment 3 in the main paper.

Results

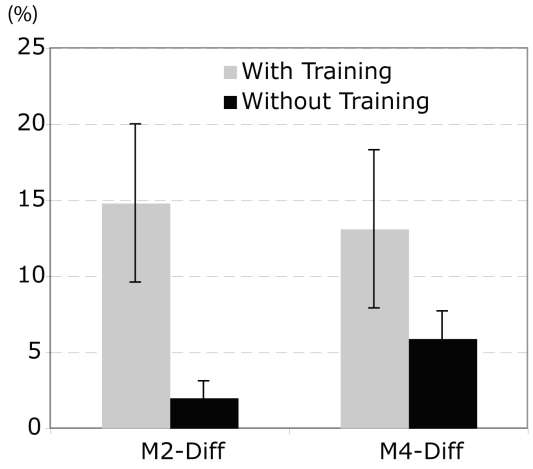
Behavioral Discrimination Performance

The data from the first and second runs of the behavioral discrimination test of the thirteen subjects in the supplementary study ("without training"), along with the data from the first and

second runs of discrimination testing of the thirteen subjects in main study ("with training"), are shown in Supplementary Figure 16, and the only significant difference between the performance of two groups was found for the second test in the M2 condition ($p<0.05$). Supplementary Figure 17 shows the difference between the first and second test for the M2 and M4 conditions for both subject groups. For both M2 and M4 conditions, improvement in the trained group is greater than for the control group by a factor of two or more. This difference is significant for the difficult M2 condition ($p<0.05$), where training-induced learning of a shape representation was expected to increase performance. For the M4 condition, while improvement was still more than twice as high for the trained than for the control group, the difference did not reach significance ($p>0.1$), possibly due to ceiling effects (average performance was at 91% correct post-training). The comparatively small increases in performance in the control group could reflect stimulus learning due to repeated exposure to the training stimuli (see Discussion in the main paper), or could reflect a task learning component (see, e.g., Furmanski and Engel, "Perceptual learning in object recognition: object specificity and size invariance," *Vision Research,* 2000 and Edelman, Bulthoff, Sklar, "Task and object learning in visual recognition", *MIT AI Memo #1348*, 1991). In summary, the data clearly show that training on the categorization task increased performance on the discrimination task and that the improvement in discrimination performance cannot be accounted for by a test/retest effect.



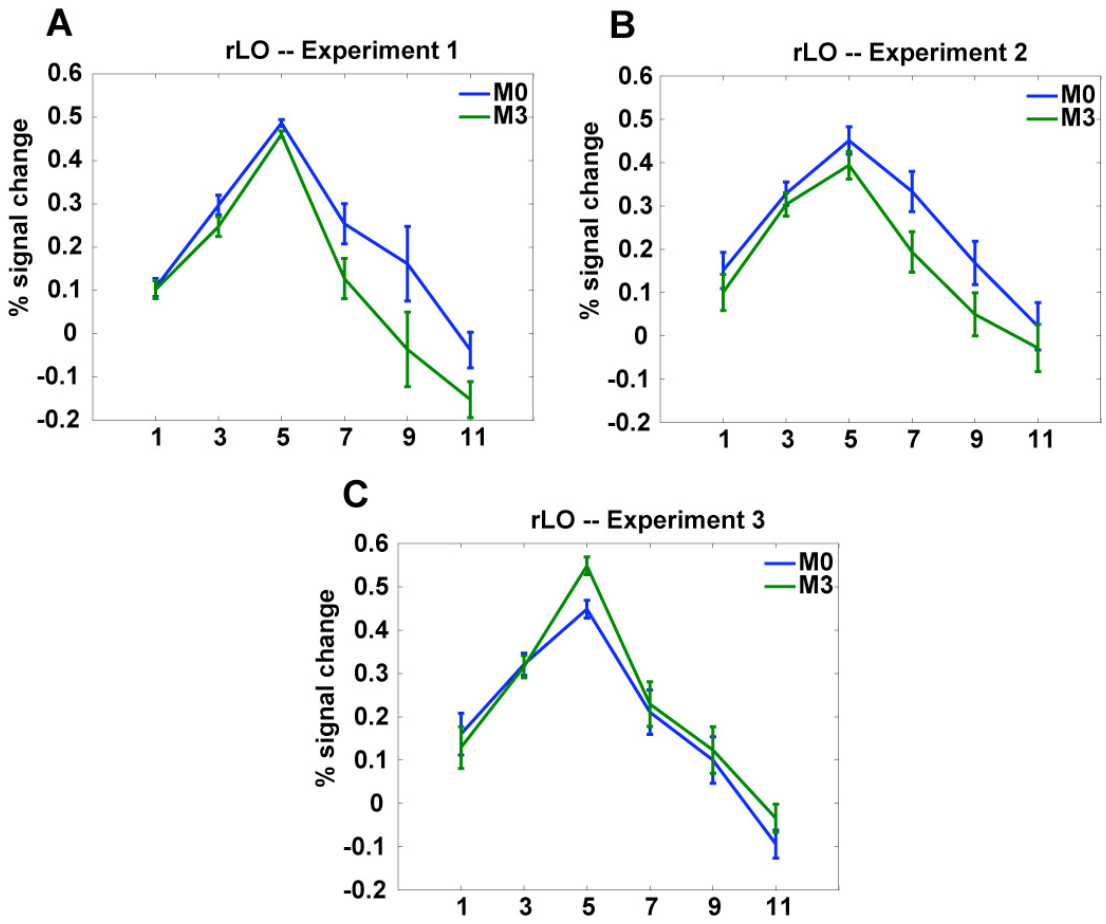
Supplementary Figure 16: *Performance of trained ("with training") and control ('without training") groups (both n=13) on the behavioral car discrimination experiment. "1st" refers to the first testing, "2nd" to the second testing session (without intervening category training for the control group, and after intervening categorization training for the trained group from the main paper). Error bars show SEM.*



Supplementary Figure 17: *Performance improvement for the two subject groups (n=13 for both). The figure shows the average difference in performance in the second relative to first discrimination testing session, for the M2 and M4 conditions. Error bars show SEM.*

fMRI Data

(1) rLO: Following the same procedures detailed in the main paper, we identified the rLO ROI for each individual subject (n=7), then extracted the fMRI responses from the ROI. For the comparison of M0 versus M3, there was no significant difference before training ($p>0.3$ for the first pre-training scan, $p>0.1$ for second pre-training scan), but a significant difference after training ($p<0.05$) (Supplementary Figure 18), suggesting that the observed increased sensitivity of the fMRI signal to shape change after categorization training could not come from a test/retest effect.



Supplementary Figure 18.. *Mean fMRI response in right LO when control subjects (n=7) were doing the same/different position task in (A) Experiment 1 (pre-training), (B) Experiment 2 (pre-training, 2nd scan), and (C) Experiment 3 (post-training). Error bars show within-subject SEM.*

(2) Whole brain analysis – identifying the brain regions involved in general same/different tasks. As the fMRI portion of this supplementary study only included 7 participants, to increase the statistical power, we treated each scan (i.e., first and second scans of pre-training, and the third scan of post-training) from each subject as from an individual subject, which resulted in a total of 21 scans. We then first identified the brain regions involved in the same/different position task with a contrast of car versus baseline (threshold $p<0.001$, uncorrected, at least 20 contiguous voxels; the same threshold was used in following voxel-wise analysis – see Supplementary

Table 5). Interestingly, none of these regions were at the same (or even close to the) location that we had identified in the main paper as the area containing category-selective neurons. Furthermore, the contrast of "different" trials versus "same" trials also failed to reveal involvement of the category-selective right lateral prefrontal cortex region from the main study (Supplementary Table 6). The results thus strongly support the notion that the category-selective right lateral prefrontal cortex region identified in the main paper was not part of a brain network involved in general Same/Different judgments, but rather was specifically activated in the categorization task (see (3) below for additional support).

(3) By contrast, when subjects were doing the categorization task (n=7), the category-selective right lateral prefrontal cortex region (highlighted in **bold** in Supplementary Table 7) was again activated by the task as indicated by the contrast of car versus baseline from the categorization scan even though there was only a rather small number of scans (n=7)[1]. It will be interesting in future studies to investigate the properties of this area in future studies using other tasks (e.g., naming) to further investigate its role in categorization, in particular whether it supports categorization across different tasks.

[1] Since we only had a small sample of subjects (n=7) for the categorization scan, we ran an additional non-parametric analysis using the SnPM toolbox (http://www.sph.umich.edu/ni-stat/SnPM/) to validate the results observed using the standard analysis methods of the SPM software package. We observed a similar activation pattern in the same regions (including left/right visual cortex, right lateral inferior frontal gyrus (z=24, the aforementioned category selective region), and a much more inferior region, right middle inferior frontal gyrus (z=-20)).

Supplementary Table 5. **Brain regions that were strongly activated by the stimuli when subjects were doing a position same/different judgment task.**

Region			**MNI Coordinates**		
	mm^3	**Z_{max}**	**X**	**Y**	**Z**
L/R Supp Motor Area	11168	6.46	-4	8	54
L Inf/Mid Occipital	30720	6.29	-12	-102	-6
		5.99	-16	-94	-10
		5.65	-30	-86	-14
R Inf Occipital/R Inf Temporal	37080	5.88	26	-98	-2
		5.61	36	-100	-2
		5.58	40	-80	-12
R Precentral/Postcentral/Inf Parietal	31512	5.59	62	-20	50
		4.92	42	-34	48
		4.88	44	-8	56
R Sup/Mid Orb Frontal	5608	5.44	18	44	-18
		4.23	48	46	-16
		4.16	34	56	-16
L Cerebellum	5984	5.04	-26	-66	-50
		4.74	-14	-60	-46
		4.16	-18	-52	-50
R Cerebellum	5032	4.89	16	-60	-46
		4.27	30	-62	-50
		3.73	16	-74	-48
L Inf Parietal/Postcentral	14504	4.62	-28	-40	40
		4.55	-42	-38	52
		4.43	-64	-26	46
R Insula/Inf Frontal	5632	4.61	34	30	-2
L Precentral	3064	4.48	-26	-8	48
		3.41	-50	-12	62
L Insula	2064	4.39	-30	22	4
L Sup/Mid Orb Frontal	552	3.99	-20	42	-18
R Calcarine	328	3.82	24	-50	10
L Rolandic	904	3.68	-44	-4	6
L Thalamus	808	3.65	-16	-16	4
R Cerebellum/Vermis	248	3.59	8	-78	-18
L Postcentral/SupraMarginal	208	3.35	-56	-20	18

Supplementary Table 6. **Brain regions that showed stronger activation to pairs of cars presented at different positions than to those presented at the same position.**

Region			**MNI Coordinates**		
	mm^3	**Z_{max}**	**X**	**Y**	**Z**
L/R Precuneus	2992	4.41	-4	-64	64
		3.53	6	-48	60
		3.42	6	-72	64
L Inf/Mid Occipital	960	4.41	-54	-78	0
L Sup/Mid Orb Frontal	2200	4.3	-24	40	-22
		4.17	-20	50	-14
		4.05	-24	58	-6
R Sup/Mid Orb Frontal	2272	4.05	20	48	-10
		4.01	20	38	-18
		3.92	24	58	-10
R Inf Temporal	536	3.91	38	-6	-42
		3.33	48	-18	-40
L Sup Frontal/Precentral	664	3.91	-32	2	64
		3.43	-20	2	68
R Inf/Mid Orb Frontal	1064	3.87	48	38	-14
R Mid Occipital	1704	3.84	34	-90	28
		3.68	38	-80	38
		3.63	44	-82	30
R Inf Parietal	968	3.83	54	-50	50
		3.35	52	-56	62
R Fusiform	328	3.71	36	-24	-30
R Precentral/Postcentral/Inf Parietal	384	3.71	26	-4	48
R Mid/Sup Temporal	304	3.66	46	-6	-18
L Inf Parietal/Postcentral	1120	3.65	-46	-54	54
		3.38	-54	-42	60
L Caudate	336	3.62	-6	6	-12
	168	3.6	0	-30	-50
L Inf Tri Frontal	160	3.58	-54	16	-2
L Inf Tri Frontal	160	3.48	-48	46	0
L Sup Med Frontal	160	3.38	-8	38	28
L Sup Temporal	168	3.33	-28	22	-28
		3.3	-34	28	-28

Supplementary Table 7. **Brain regions that were strongly activated by the stimuli when subjects were doing the category same/different judgment task.**

Region			**MNI Coordinates**		
	mm^3	**Z_{max}**	**X**	**Y**	**Z**
R Inf Occipital/R Inf Temporal	8824	5.2	30	-70	-26
		4.64	28	-42	-20
		4.42	26	-50	-22
R Inf Frontal Gyrus	**288**	**4.74**	**52**	**24**	**26**
R Mid Orb Frontal	232	4.73	24	38	-20
L Inf/Mid Occipital	4504	4.33	-18	-94	-14
		4.31	-14	-100	-6
		4.19	-14	-94	4
R Inf Oper Frontal	792	4.31	48	6	24
L Cerebellum	744	4.03	-26	-60	-50
		3.78	-34	-60	-48
L Supramarginal	200	3.87	-42	-42	32
L Precentral	176	3.79	-34	-12	52
		3.42	-40	-16	56
R Mid Frontal	168	3.77	48	40	20
R Sup Parietal	600	3.68	32	-62	56
L/R Supp Motor Area	840	3.68	0	8	56