Published in final edited form as:
Neuron. 2007 March 15; 53(6): 891–903.

Categorization Training Results in Shape- and Category-Selective Human Neural Plasticity

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*

Summary

Object category learning is a fundamental ability, requiring combination of “bottom-up” stimulus-driven with “top-down” task-specific information. It therefore may be a fruitful domain for study of the general neural mechanisms underlying cortical plasticity. A simple model predicts that category learning involves the formation of a task-independent shape-selective representation that provides input to circuits learning the categorization task, with the computationally appealing prediction of facilitated learning of additional, novel tasks over the same stimuli. Using fMRI rapid-adaptation techniques, we find that categorization training (on morphed “cars”) induced a significant release from adaptation for small shape changes in lateral occipital cortex irrespective of category membership, compatible with the sharpening of a representation coding for physical appearance. In contrast, an area in lateral prefrontal cortex, selectively activated during categorization, showed sensitivity post-training to explicit changes in category membership. Further supporting the model, categorization training also improved discrimination performance on the trained stimuli.

Introduction

Object category learning is a fundamental cognitive ability essential for survival, as exemplified by the obvious importance of efficiently distinguishing friend from foe or edible from poisonous objects. Category learning is also a convenient and rich domain in which to study the general neural mechanisms underlying cortical plasticity, as it requires combining “bottom-up” stimulus-driven information with “top-down” task-specific information. Recent monkey studies (Freedman et al., 2003;Op de Beeck et al., 2001;Thomas et al., 2001) have provided support for a two-stage model of perceptual category learning (Ashby and Spiering, 2004;Nosofsky, 1986;Riesenhuber and Poggio, 2000;Sigala, 2004;Thomas et al., 2001), involving a perceptual learning stage in extrastriate visual cortex in which neurons come to acquire sharper tuning with a concomitant higher degree of selectivity for the training stimuli. These stimulus-selective neurons provide input to task modules located in higher cortical areas, such as prefrontal cortex (Freedman et al., 2003), that can then learn to identify, discriminate, or categorize the stimuli. A computationally appealing property of this hierarchical model is that the high-level perceptual representation in visual cortex can be used in support of other

Correspondence should be addressed to M.R. (mr287@georgetown.edu). Department of Neuroscience, Georgetown University Medical Center, Research Building Room WP-12, 3970 Reservoir Rd. NW, Washington, DC 20007, USA

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

tasks involving the same stimuli (Riesenhuber and Poggio, 2002), permitting transfer of learning to novel tasks. For instance, a population of neurons tuned to views of different cats and dogs (Freedman et al., 2003) could provide input to a classifier discriminating cats from dogs, as well as also allowing either the identification of a specific dog ("my dog Rosie") or its categorization at a different level ("black Labrador retriever").

While not possessing the temporal and spatial resolution of single unit recording studies, functional neuroimaging studies of category learning offer distinct advantages, including the ability to directly study complex task training effects in humans in a before/after comparison, sampling the entire brain, whereas physiology studies are usually limited to recording from just one or two brain regions and have to rely on indirect comparisons to estimate learning effects, perhaps by comparing neuronal selectivities for trained and untrained stimulus sets in the same animal (Freedman et al., 2003).

Neuroimaging studies of learning commonly compare blood oxygenation-level dependent (BOLD)-contrast responses to objects before and after training. However, given that total neuronal activity in a voxel containing hundreds of thousands of neurons depends on the number of active neurons as well as their selectivity, learning-induced sharpening of neural responses – which by itself would lead to a lower population response as each neuron responds to fewer stimuli (Freedman et al., 2006;Rainer and Miller, 2000) – could lead to either decreases or increases in neuronal activity, depending on how training affects the number of selective neurons. This makes it difficult to interpret BOLD-contrast amplitude changes as a measure of tuning selectivity. Indeed, previous functional magnetic resonance imaging studies (fMRI) studies have found that perceptual and category learning can induce BOLD-contrast signal response increases (Gauthier et al., 1999;Op de Beeck et al., 2006;Pollmann and Maertens, 2005), decreases (Reber et al., 1998), or both (Aizenstein et al., 2000;Kourtzi et al., 2005;Little and Thulborn, 2005).

To more directly probe the changes in neuronal tuning resulting from category acquisition, we trained a group of human participants to categorize stimuli ("cars") generated by a morphing system that was capable of finely and parametrically manipulating stimulus shape (Shelton, 2000), a technique employed in our earlier monkey studies of category learning (Freedman et al., 2003). This approach allowed us to precisely define the categories and dissociate category selectivity, which requires neurons to respond similarly to dissimilar stimuli from the same category as well as respond differently to similar stimuli belonging to different categories (Freedman et al., 2003), from mere tuning to physical shape differences, where neuronal responses are a function of physical shape dissimilarity, without the sharp transition at the category boundary that is a hallmark of perceptual categorization. Importantly, unlike earlier studies, we recorded brain activation before and after training using fMRI rapid adaptation (fMRI-RA) techniques, which can probe neuronal selectivity more directly than can conventional methods relying on average BOLD-contrast stimulus responses (Gilaie-Dotan and Malach, 2007;Grill-Spector et al., 2006;Jiang et al., 2006;Kourtzi and Kanwisher, 2001).

We provide direct evidence that training on a perceptual categorization task leads to the sharpening of stimulus representation coding in lateral occipital cortex (LO), a part of the lateral occipital complex (LOC) postulated to play a key role in human object recognition as the human homologue of monkey area IT (Grill-Spector, 2003;Grill-Spector et al., 2001;Kourtzi and Kanwisher, 2001). While this LO representation showed no explicit category selectivity, seeming to be selective for physical stimulus shape only, an area in the right lateral prefrontal cortex (rLPFC) exhibited category-selective responses. When participants were judging the category membership of cars, this activity was modulated by explicit changes of category membership, but not by shape differences alone. This category-selectivity was not detectible when participants were doing a position displacement task with the same stimuli, suggesting that these category circuits were only active when categorization was an explicit component of the task. Furthermore, we found that categorization training also improved subject performance on a discrimination task involving the car stimuli, without additional training. These observations provide strong support for the aforementioned model of perceptual categorization which posits that category learning involves two components: the learning of a shape-sensitive but task-independent representation that provides input to circuits responsible for categorization. Finally, the results show that fMRI-RA techniques can be used to investigate learning effects at a more direct level than conventional approaches based on comparing average BOLD-contrast response amplitude in response to individual conditions, providing a powerful new tool to study the mechanisms of human cortical plasticity.

Results

Behavior

Participants were trained to categorize a continuous set of stimuli that spanned two categories, each based on two different car prototypes (Figure 1). The morphed images were linear combinations of all possible arrangements between prototypes. By blending differing prototype amounts from the two categories, we could continuously vary the object shape and precisely define the category boundary. After an average of 5.25 (±0.48) hours of training, participants were able to judge the membership of the morphed cars reliably (see Methods Section and Figure 2).

fMRI Experiments 1 and 2 (Displacement Detection Task)

The first prediction of our two-stage model of category learning is that categorization training leads to sharper neuronal shape selectivity to trained car images in extrastriate visual cortex. To explore changes in neuronal shape selectivity using fMRI, we adopted an event-related fMRI-RA paradigm (Jiang et al., 2006;Kourtzi and Kanwisher, 2001), in which a pair of car images of varying shape similarity was presented in each trial. The fMRI-RA approach is motivated by findings from IT monkey electrophysiology experiments that showed that when pairs of stimuli were presented sequentially, a smaller neural response was observed following presentation of the second stimulus (Lueschow et al., 1994;Miller et al., 1993). It has been suggested that the degree of adaptation depends on stimulus similarity, with repetitions of the same stimulus causing the greatest suppression. In the fMRI version of this experiment, the BOLD-contrast response to a pair of stimuli presented in rapid succession was measured for pairs differing in specific perceptual aspects (*e.g.,* viewpoint or shape), and the combined response level was assumed to predict stimulus representational dissimilarity at the neural level (Grill-Spector et al., 2006;Murray and Wojciulik, 2004). Indeed, we (Jiang et al., 2006) and others (Fang et al., 2006;Gilaie-Dotan and Malach, 2007;Murray and Wojciulik, 2004) have recently provided evidence that parametric variations in shape, orientation, or viewpoint – stimulus parameters putatively associated with neuronal tuning properties in specific brain areas – are reflected in systematic modulations of the BOLD-contrast response, suggesting that fMRI adaptation could be used as an indirect measure of neural population tuning (Grill-Spector et al., 2006). Following this hypothesis, we reasoned that if categorization training leads to sharpened neuronal selectivity to car images, then the overlap of neuronal activations caused by two sequentially presented car images differing by a fixed amount of shape change would decrease following training, resulting in an increase of BOLD contrast response in the car-selective regions.

Previous studies (Grill-Spector et al., 2001;Kourtzi and Kanwisher, 2001;Kourtzi et al., 2003;Murray and Wojciulik, 2004) have suggested that LOC plays a central role in human object recognition and we therefore hypothesized that training-induced learning effects should occur in this area. LOC consists of two sub-regions, LO (lateral occipital) and pFs (posterior fusiform). In this study, we focused on the LO region, as the pFs region could not be reliably identified by our localizer paradigm in about half of the participants. To probe training effects on LO neurons, we scanned participants before and after training using an event-related RA paradigm with a displacement detection task for which categorization training was irrelevant, thus avoiding potentially confounding influences due to the change of task difficulty as a matter of training (Gerlach et al., 1999) and other potential confounds caused by top-down effects of the task itself (Freedman et al., 2003;Grady et al., 1996;Sunaert et al., 2000).

Stimulus pairs of controlled physical dissimilarity were created with the morphing system. In particular, we created pairs of identical images (condition M0), and pairs of images differing by 33.33% shape change, with both cars in a pair either belonging to the same category, $M3_{within}$, or to different categories, $M3_{between}$ (Figure 3A). This made it possible to attribute possible signal differences between $M3_{within}$ and $M3_{between}$ to an explicit representation of the learned categories. The regions of interest (ROI) were identified independently for each subject using localizer scans (see Methods). We then extracted the BOLD-contrast time series from these independently identified ROI. Since the fMRI response at the right LO (rLO) peaked at the time window of 4-6s after the onset of each trial, statistical analyses (repeated measures ANOVA followed by planned t-tests) were carried out on the peak BOLD-contrast values. Before categorization training (Experiment 1), there were no significant differences across the three conditions (M0, $M3_{within}$, and $M3_{between}$), $p>0.3$ (Figure 3B, left). Additional paired t-tests between M0 and the mean of $M3_{within}$ and $M3_{between}$ also showed no difference ($p>0.5$). This indifferent response suggests that the neuronal responses to car images in rLO in the experiment showed little sensitivity to cars differing in shape by 33.33%. By contrast, after categorization training (Experiment 2), a significant difference was observed across the three conditions using the same paradigm and stimuli, $F(2, 32)=5.219$, $p=0.014$ (Figure 3B, right). Post-hoc t-tests revealed significant differences between M0 and $M3_{within}$ ($p<0.05$), and between M0 and $M3_{between}$ ($p<0.05$), but not between $M3_{within}$ and $M3_{between}$ ($p>0.4$). Additionally, for the data from the fifteen participants whose data were included in both data sets (pre- and post-training), a repeated measures ANOVA revealed 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 ($p>0.1$) (Figure S5). A control study showed that these effects could not be accounted for by test-retest effects but rather appeared to be due to the intervening category training (see Supplementary Material). Taken together, these data indicate that, after training, neurons in the rLO ROI showed a greater response difference to the same stimulus shape difference when compared to the period before training, suggesting that categorization training sharpened the tuning of LO neurons to the car stimuli. Furthermore, the non-differential response at LO between the $M3_{within}$ and $M3_{between}$ conditions suggested that LO neuron tuning was largely determined by stimulus shape and not category membership.

One interesting prediction of the two-stage model of category learning is that the high-level shape-based representation learned as a result of categorization training can also be recruited for different tasks on the same stimuli, *e.g.*, to support improved discrimination of these stimuli relative to untrained participants. Indeed, we found that categorization training also improved participants' performance on a car discrimination task (Figure 4). Crucially, this improvement was not limited to sections of the stimulus space relevant for categorization (*i.e.*, the boundary region between the two categories), but was also found away from the boundary and, most importantly, for within-category morph lines, as would be predicted for a “bottom-up” shape-based neural representation of car shape. A control study showed that this improvement in behavioral discrimination performance could not be accounted for by a test/retest effect on the discrimination task (see Supplementary Material).

In contrast, it has been suggested that the FFA mediates the subordinate-level discrimination of objects of expertise (Gauthier et al., 1999). We therefore tested whether categorization training also led to increased sensitivity to shape changes in the FFA. Interestingly, in contrast to LO, a repeated-measures ANOVA of the peak values in the right FFA (rFFA) revealed no difference among the three conditions before (Experiment 1, $p>0.3$), or after training (Experiment 2, $p>0.4$, Figure 3C). This finding suggests that the selectivity of FFA neurons was not affected by category training, and that the improvement in discrimination ability for the trained objects was more likely to be mediated by the increased car shape sensitivity of LO neurons, as predicted by recent modeling studies (Jiang et al., 2006).

The data from left LO and FFA did not show significant selectivity to the 33.33% shape change of car images either before or after training (Figure S7). We also did not find any differential activation among the three conditions in early visual cortex (see Methods), either before ($p>0.4$) or after training ($p>0.2$) (Figure S8), suggesting that the observed learning effects were unlikely to be non-specific or global phenomena.

For both Experiments 1 and 2, we examined possible changes in other brain regions by conducting a voxel-wise whole-brain analysis (see Methods) using contrasts of $M3_{between}$ > $M3_{within}$ and M0 to detect category-selective brain regions, and $M3_{between}$ and $M3_{within}$ > M0 to detect any shape-selective brain regions. These contrasts did not reveal any brain regions of at least 20 contiguous voxels at a threshold of $p<0.001$ (uncorrected).

fMRI Experiment 3 (Categorization Task)

To probe which brain regions exhibited category-related activations, and thus might include category-selective neurons, we scanned our participants again post-training using the same fMRI-RA paradigm, this time while they were performing a categorization task requiring them to judge whether the two cars shown in each trial belonged to the same or different categories. In addition to three conditions tested in Experiment 1 and 2, a fourth condition (M6) was added, with the two cars in each M6 trial belonging to different categories, with 66.67% shape change between them (Figure 5A). Thus, the pair of cars of M0 and $M3_{within}$ belonged to the same category, while the pairs of cars of $M3_{between}$ and M6 belonged to different categories. We predicted that brain regions containing category-selective neurons should show stronger activations to the $M3_{between}$ and M6 trials than to the $M3_{within}$ and M0 trials, as the stimuli in each pair in the former two conditions should activate different neuronal populations while they would activate the same group of neurons in the latter two conditions.

As in Experiments 1 and 2, statistical analyses were first carried out on the peak of the fMRI responses at the independently defined ROI. As the peak of fMRI response in the rLO regions lasted more than one TR (3rd and 4th TR after the onset of each trial), statistical analysis was carried out on the mean of 3rd and 4th TR (Figure 5B). Repeated measures ANOVA revealed significant differences among the four conditions (M0, $M3_{within}$, $M3_{between}$, and M6), $F(3, 45) = 8.515$, $p=0.001$. Post-hoc paired t-tests revealed a significant difference between M0 and $M3_{within}$ ($p=0.01$), between M0 and $M3_{between}$ ($p<0.0005$), between M0 and M6 ($p < 0.00005$), between $M3_{between}$ and M6 ($p<0.05$), but not between $M3_{within}$ and M6 ($p>0.15$) or between $M3_{within}$ and $M3_{between}$ ($p>0.9$). The effects in rLO not only confirmed the findings of Experiment 2, in which a car with 33.33% shape change already appeared to activate a substantially different populations of rLO neurons, but also suggested that there was still substantial overlap between the population of rLO neurons responding to a particular car and those responding to a car with a 33.33% relative shape change (Jiang et al., 2006), as indicated by the significant difference between $M3_{between}$ and M6. Thus, there was no evidence for category selectivity in rLO even while participants were performing the categorization task.

In contrast to the car shape selectivity in rLO, no significant difference was found in early visual cortex (Figure S10), nor at the right FFA, F(3, 45)=1.709, p=0.20 (Figure 5C). However, since the voxel-wise analysis revealed a cluster of voxels in the fusiform gyrus showing a significantly stronger response on M6 trials than on M0 trials (Figure S11), we conducted additional paired t-tests, and found a significant difference between M0 and M6 (p=0.01), but not for any other comparisons. This difference between M0 and M6 could either be due to the involvement of the face-selective FFA when viewing trained objects (Gauthier et al., 1999), or it could be due to the overlap between the face-selective FFA and nearby object-selective pFs regions (Grill-Spector et al., 2004). Since we could not reliably identify the pFs region in this study as mentioned earlier, to test these two hypotheses directly, we redefined two new ROI, a “core’“ face ROI and a “surround” face ROI in the fusiform gyrus for each individual subject (see Methods). The voxels in the former responded more strongly to faces than those of the latter (Figure S12). We then extracted the BOLD-contrast response in the two newly defined ROI from the event-related scans (Figures 5D and 5E). An ANOVA with two ROI and four conditions as repeated measures revealed that peak BOLD responses to car images in the “core” face ROI were significantly higher than those in the “surround” face ROI, F(1,15) =7.326, p<0.05, likely because the “surround” face ROI included regions anterior to the “core” face ROI which are not part of pFs. More importantly, there was a significant interaction between the ROIs and the four conditions, F(3, 45)=3.194, p<0.05, and a marginal effect among the four similarity conditions, F(3, 45)=2.293, p=0.12. The significant interaction indicated that the difference among the four conditions was stronger in the “surround” face than in the “core” face ROI. We then conducted an additional ANOVA with four conditions as repeated measures on the two sets of data separately, and a significant difference was found in the “surround” face ROI, F(3, 45) = 3.274, p<0.05, but not in the “core” face ROI, F(3, 45) = 1.510, p>0.2. The data thus demonstrated that the differences among the four conditions were stronger in the “surround” face ROI than that in the “core” face ROI, suggesting that the difference in the FFA ROI was less likely caused by the differential response of face selective neurons in the FFA, but rather more likely due to an overlapping with nearby pFs regions, which has been shown to exhibit strong repetition-suppression for non-face objects (Grill-Spector et al., 1999). The data from left LO and FFA are shown in Figure S9.

We then conducted a whole brain analysis (see Methods) to examine the brain regions that were involved in the categorization task. The brain regions significantly activated in the categorization task versus baseline included the visual cortex, motor cortex, frontal cortex, parietal cortex, insular cortex, and the thalamus (Table S1 in the Supplementary Material). To probe the brain regions that were sensitive to category differences, we first compared the activation of M6 versus M0 since participants could very reliably judge the category memberships of the pair of cars in the M0 and M6 conditions. As listed in Table 1, many brain regions, including prefrontal, parietal, and inferior temporal cortices showed stronger activations to M6 than to M0, further supporting the involvement of these brain areas in the representation of learned stimulus categories (see also Moore et al., 2006). To further examine the differential activations to trials in which the two cars belonged to the same ($M3_{within}$ and M0) versus different categories (M6 and $M3_{between}$), a comparison of M6 and $M3_{between}$ versus $M3_{within}$ and M0 was conducted, and similar brain regions were found (Table 1). This selectivity was not due to reaction time differences in the different conditions (Tables S2 and S3).

While both the comparisons of M6 versus M0, and M6 and $M3_{between}$ versus $M3_{within}$ and M0 revealed that the PFC, parietal, and inferior temporal regions showed stronger activation when the two cars belonged to different categories than when they belonged to the same category, the inclusion of the M0 and M6 conditions to investigate *category* tuning (i.e., unconfounded by tuning to mere differences in physical shape) suffers from a confound due to the different amounts of shape change in the M0 and M6 conditions. By contrast, the comparison of

$M3_{between}$ versus $M3_{within}$ represents the most direct comparison for category-related activity, as the stimulus pairs in both conditions differed by the same relative amount of shape change, but either crossed or did not cross the category boundary, respectively. However, these conditions required participants to determine the category memberships of stimuli close to the category boundary, making these conditions particularly difficult and susceptible to small variations in participants' individual category boundaries for the different morph lines (see Figures 2 and S2), in particular for the $M3_{between}$ condition, which required comparing the category memberships of two stimuli close to the category boundary. Indeed, the comparison of $M3_{between}$ versus $M3_{within}$ across all four morph lines was not sensitive enough to identify category-selective brain areas. For a more sensitive analysis, we re-modeled the fMRI response with a 4×4 setup (consisting of the 4 above-mentioned conditions × 4 morph lines). We then identified, for each subject individually, the morph line on which participants had the highest behavioral performance inside the scanner (Figure S13), and probed category-related brain regions with the contrast of $M3_{between}$ versus $M3_{within}$ for this "best" morph line only, predicting that high behavioral performance on these conditions would result from neurons sharply tuned to the different categories and thus produce a higher signal difference between $M3_{between}$ and $M3_{within}$. Interestingly, the only region that showed stronger activation using this contrast was in the rLPFC, at a location similar to that found in the previous comparisons (see Table 2 and Figure 6A and B). The comparisons of M6 versus M0, and of M6 and $M3_{between}$ versus $M3_{within}$ and M0 on the "best" morph line also found the same rLPFC region (see Tables 1 and 2). Thus, the most striking and consistent finding when comparing the category-selective activations was that the same rLPFC region was found to be activated more strongly when the two cars belonged to different categories than when the two cars belonged to the same category under all comparisons. Note that this differential activation could not be explained by task-related motor responses, which were counter-balanced across participants.

To test the predicted mechanistic relationship between rLPFC activation and categorization performance, we went back to the ROI defined by M6 > M0 on the "best" morph line (see Table 2 and Figure 6A) and examined the correlation of the difference between the fMRI response for the $M3_{between}$ and $M3_{within}$ conditions in this ROI (as an index of how sharply neurons in this area differentiated between the two categories; note that the ROI definition, M6 versus M0, was independent of the conditions involved in the correlation analysis, $M3_{between}$ versus $M3_{within}$) and the average of the behavioral categorization accuracy on those trials within the scanner (as a measure of behavioral performance), predicting a positive correlation between the two variables. Of special concern for this analysis is the fact that low performance on those conditions could either be due to weak category tuning of neurons (the effect of interest, with the predicted effect of a positive correlation between fMRI activation and behavior) or due to subject inattentiveness or failure to perform the task in the scanner (in which case we would not expect a similarly tight relationship between fMRI and behavior). Indeed, calculating the correlation between fMRI activation and behavior over all participants only produced a marginal correlation (r=0.206, p=0.102, Figure 6E). To focus on the participants who were most likely to have consistently performed the task in the scanner, we performed a second correlation analysis, excluding participants with an average performance on the easy M0 and M6 conditions below 85% correct. As predicted, the remaining 11 participants showed a high degree of correlation between "same category" ($M3_{within}$) vs. "different category" ($M3_{between}$) activation difference in the category ROI and behavioral performance (r=0.409, p<0.01, Figure 6F). Such a correlation strongly supports a key role of right lateral PFC in object categorization, in particular that rLPFC contains neurons sharply tuned to different categories, with the degree of category selectivity determining the behavioral performance. This causal role of rLPFC in determining participants' categorization decision is also reflected by a significant modulation of activation in this area with participants' "same/different category" responses in the M3 conditions (Figure S14). Notably, this brain region (rLPFC) was not active when participants performed a "same/different position" task on the stimuli (see Supplementary Material), suggesting that activation in this area was indeed specific to the categorization component of the task in Experiment 3, and did not reflect generic "same/different" processing.

Finally, based on previous studies (Vogels et al., 2002), category-related activation in PFC would be expected to be much weaker, or even abolished for the same stimuli if participants were doing a task for which the learned categories were irrelevant, e.g., the displacement detection task of Experiment 2. To test this hypothesis, we extracted the signal change in Experiment 2 at the categorization ROI based on the $M3_{between}$ vs. $M3_{within}$ contrast on the "best" morph line (see Table 2 and Figure 6B). For the data collapsed across morph lines, no difference was found among the three conditions (M0, $M3_{within}$, and $M3_{between}$) in Experiment 2. For a more sensitive analysis, as in Experiment 3, we performed an ANOVA on the trials with stimuli from the same morph line on which each individual subject had the best performance in Experiment 3 (Figure 6C). No significant difference ($p>0.5$) was found among the three conditions (M0, $M3_{within}$, and $M3_{between}$) in Experiment 2. Additional paired t-tests also found no difference between M0 and $M3_{between}$, between $M3_{within}$ and $M3_{between}$, and between $M3_{between}$ and the mean of M0 and $M3_{within}$. Similar results were obtained when the ROI was defined by the comparison of M6 versus M0, or M6 and $M3_{between}$ versus $M3_{within}$ and M0 (see Figure S15). In summary, strong category-selective activation in rLPFC was found only when participants were explicitly doing a categorization judgment task, suggesting that the category-selective circuits learned as a result of training were only active when the subject was performing the corresponding categorization task.

Discussion

Previous monkey electrophysiology studies have suggested that perceptual learning in object recognition tasks could sharpen the tuning of neurons in inferotemporal cortex (Freedman et al., 2006), and recent theoretical work has suggested that similar mechanisms might play a role in human object discrimination (Jiang et al., 2006). In our study, we used an fMRI rapid adaptation paradigm designed to probe neuronal tuning more directly than previous studies of human perceptual learning that focused on the average BOLD-contrast response to the training stimuli. Testing the same participants before and after training, we found that while pre-training, there was no indication of selectivity of neurons in LO for the target stimuli (as response levels in the adaptation experiment did not differ between the M0 and M3 conditions), training on a perceptual categorization task involving fine discriminations among the target objects led to a release from adaptation in fMRI for small shape changes (M3 vs. M0) post-training, compatible with the notion that LO neurons acquired increased selectivity for the training stimuli through training.

Our failure to find evidence for the sharpening of neuronal tuning in the FFA region (see also Yue et al., 2006) despite the significant improvement of participants' discrimination abilities for the training class in general (and not just at the category boundary) is in line with the two-stage model of category learning that predicts that category training leads to the learning of a shape-specific representation dedicated to the object class of interest (i.e., disjoint from the face-tuned neurons in the FFA (Jiang et al., 2006)) that can provide input to circuits learning different tasks, such as categorization or discrimination, and thus permit transfer of learning from one task to another (Jiang et al., 2006;Riesenhuber and Poggio, 2002). The data are more difficult to reconcile with proposals (Tarr and Gauthier, 2000) that have postulated that the FFA serves to learn and mediate the discrimination of objects of expertise in general (i.e., not just faces). In particular, unlike the results for LO, we did not find any differential activation (between the M0 and M3 conditions) in the FFA as a result of training when participants were doing the position displacement task, despite an improvement in participants' abilities to discriminate the stimuli and despite similar amounts of training as in earlier studies (Gauthier et al., 1999) that have reported training effects in the FFA. Differential activation was found for the M0 and M6 conditions in Experiment 3, and group analysis also showed a region in the fusiform area with significantly higher response in the M6 vs. the M0 condition. However, it appeared that the selectivity observed in the fusiform region was more likely due to a spatial overlap between the object-selective pFs region and the face-selective FFA (Grill-Spector and Malach, 2004, see also Rhodes et al., 2004), rather than due to a car-selectivity of the face-selective neurons per se, as (1) the ROI-based analysis in the FFA showed a smaller difference than the whole-brain based analysis; and (2) the "core FFA" that included highly face-selective voxels showed smaller differential activity for the different conditions than the nearby regions that included less face-selective voxels. (see Figure S12 for additional analyses and support).

The prefrontal cortex is generally assumed to play a key role in categorization. Our previous monkey studies (Freedman et al., 2003), using a very similar categorization task, have shown that after categorization training, some neurons in PFC come to be category-selective, responding similarly to exemplars from one category and showing lower responses to exemplars from the other category. Using an fMRI-RA paradigm, we here provide evidence that category training similarly can lead to the learning of a population of category–selective neurons in human lateral PFC (mainly in the right inferior frontal gyrus), whose category-selectivity can be dissociated from mere shape selectivity. Furthermore, we found that the same region failed to show significant category-selective activation when participants were doing a task unrelated to categorization, similar to earlier studies (Vogels et al., 2002), in line with a role of PFC as the center of cognitive control (Miller and Cohen, 2001) that contains different task-specific circuits whose activations depend on the subject's goals.

Our data therefore support a model of perceptual categorization in which a neural representation selective for the shapes of the target objects located in LOC (or IT, in monkeys) provides input to category-selective circuits in prefrontal cortex. Importantly, the model posits that the learning of the shape-selective representation can proceed in an unsupervised fashion, driven by bottom-up stimulus information (i.e., shape) (Riesenhuber and Poggio, 2000). Such a learning scheme is both computationally simple and powerful (Serre et al., 2007). Further supporting this model, we have recently shown (Freedman et al., 2006) that even passive viewing of training stimuli can induce sharpening of IT responses to these stimuli. In contrast, a previous monkey physiology study (Sigala and Logothetis, 2002) has reported increased selectivity for category-relevant over category-irrelevant features in IT following category training. While our fMRI experiment did not include within-category morph line conditions that could be compared against the responses for the cross-category morph lines, our behavioral data that found no difference in discrimination performance on within- and cross-category morph lines argue against an underlying shape representation differentially sensitive for category-relevant and -irrelevant features in our case, in line with other monkey physiology studies in IT (Op de Beeck et al., 2001). It will be interesting to investigate this question in further studies. An intriguing possibility is that top-down feedback from prefrontal cortex may induce category-specific modulations of IT neuron activity under certain task conditions (Freedman et al., 2003, see also Miyashita et al., 1998).

While we did not find strong category selectivity in the basal ganglia, a number of studies have suggested that the basal ganglia are also involved in human category learning (Ashby and Spiering, 2004). This difference might just be trivially due to the possibility that category-related signals in the basal ganglia were not strong enough to be significant in our study. However, given that we only imaged participants after they had fully learned the task, an alternative explanation could be that the basal ganglia show stronger activity early in category learning that is reduced as participants become proficient at the task, as suggested by a recent fMRI study (Little et al., 2006). Finally, the differences might be due to the fact that the learning of different types of categorization tasks depends on multiple neural systems (Ashby and

Spiering, 2004), with the basal ganglia playing a stronger role in rule-based and information integration-based categorization, rather than the perceptual categorization studied here.

The right LPFC region showed the strongest sensitivity to change of category membership in this study. Several other regions, such as parietal cortex, occipital temporal regions, and other parts of frontal cortex were also strongly activated during the categorization task, and showed stronger activations in the M6 than in the M0 conditions. Interestingly, however, the activity in these regions did not reach significance for the stricter $M3_{between}$ vs. $M3_{within}$ comparison that dissociated shape from category tuning. Given that other recent studies have suggested that these regions might be also involved in category learning (Freedman and Assad, 2006) and expertise effects (Moore et al., 2006), the future investigation of the differential roles of these areas in category learning is of particular interest to understand how bottom-up and top-down information interact in the brain to permit the learning of novel tasks.

Methods

Participants

Twenty-two (13 female, aged 19-27) normal right-handed members of the Georgetown University community participated in this study. Experimental procedures were approved by Georgetown University's Institutional Review Board, and written informed consent was obtained from all participants prior to the experiment. Two participants participated in fMRI Experiment 1 only since they failed to reach criterion in the category learning task, thus their data were discarded. All other twenty participants participated in fMRI Experiments 1 and 2, and 17 of them participated in fMRI Experiment 3. Because of excessive head motion, the data from three participants (Experiment 1), two participants (Experiment 2), and one subject (Experiment 3) were excluded from further analysis.

Visual stimuli

A large continuous set of images was generated from four car prototypes (Figure 1A) using a 3D shape morphing algorithm (Shelton, 2000) that we have used previously to study categorization learning in monkeys (Freedman et al., 2003). The algorithm finds corresponding points between one of the prototypes and the others and then computes their differences as vectors. Morphs were created by linear combinations of these vectors added to that prototype. For more information see http://www.cs.ucr.edu/~cshelton/corr/. By morphing different amounts of the prototypes we could generate thousands of unique images, continuously vary shape, and precisely define a category boundary (Figure 1B). The category of a stimulus was defined by whichever category contributed more (>50%) to a given morph. Thus, stimuli that were close to, but on opposite sides of, the boundary could be similar, whereas stimuli that belonged to the same category could be dissimilar. This careful control of physical similarity within and across categories allowed us to disentangle the neural signals explicitly representing category membership from those related to physical stimulus shape. The four prototype stimuli were chosen from an initial set of 15 based on pilot experiments that showed that these four prototypes were of comparable perceptual dissimilarity. The stimuli differed along multiple feature dimensions and were smoothly morphed, i.e., without the sudden appearance or disappearance of any feature. They were grayscale images, 200×200 pixels in size with identical shading.

Categorization training and testing

Using a two alternative forced choice (2AFC) paradigm (Figure S1), the participants, who had no prior knowledge about the definitions of the two categories, were trained to categorize images chosen from the car morph space. Each trial started with a 300ms fixation period, which was followed by three sequentially presented car images, each presented for 400ms and separated by a 300ms static random mask. The participants' task was to judge whether the second or the third car belonged to the same category as the first car by pressing one of two buttons. Auditory feedback was given to subject on incorrect trials, and the next trial would start 800ms after participants' response or 2300ms after the offset of the third car if participants failed to make a response. Following a similar training procedure as in our previous monkey studies (Freedman et al., 2003), participants were first trained to categorize the prototype cars (containing 0% morphs from prototypes belonging to the other category). We then gradually increased the difficulty of the categorization task by introducing morphs with increasingly greater contributions from the other category until participants could reliably (performance > 80%) identify the categorical membership of randomly chosen cars that consisted of up to 40% of prototypes from the other category. Participants were trained at the pace of one hour per weekday in a continuous manner with a maximum of two weeks. On average, participants completed the most difficult level after 5.25 (±0.48) hours of training.

Stimuli were presented to participants on an LCD monitor on a dark background, at a resolution of 1024×768 with 60 Hz refresh rate, at a distance of 60cm. A customized version of Psychtoolbox (Pelli, 1997) running under MATLAB (The Mathworks, MA) was used to present the stimuli and to record the responses.

After participants reached the highest level of task difficulty, their categorization performance along the four morph lines was measured at a morph step discretization of 20 steps (in increments of 5% morph difference) between the two prototypes using the same 2AFC paradigm as in the training period but without feedback. Note that different cars were used during training (where images were randomly chosen from the morph space) and testing (where images were constrained to lie on the relevant morph lines).

Discrimination testing

To study whether categorization training also led to improvements of participants' ability to discriminate car images in general, thirteen out of twenty participants were tested on a shape discrimination task involving pairs of cars chosen along the six morph lines using the same 2AFC paradigm as described above, both before and after categorization training. To ensure subject performance was above chance even before any training, match/non-match shape differences of 20% (M2) and 40% (M4) were tested in different trials. Stimuli were chosen from all six morph lines (including four cross-category and two within-category morph lines, see Figure 1) discretized in steps of 20% shape change, as in the example morph line of Figure 1B. This resulted in ten unique trials for each morph line (six pairs with 20% difference and four pairs with 40% difference). Each trial was repeated 12 times, for a total of 120 trials per morph line and a grand total of 720 trials tested pre- as well as post-training.

Functional localizer scans

Using a block design, the EPI images from two functional localizer scans were collected to define the car-selective regions in the lateral occipital cortex (LO) and the face-selective regions in the fusiform gyrus – one at the beginning of each session and one at the end. During each localizer run, following an initial 10s fixation period, 50 grayscale images of cars, scrambled cars, and faces were presented to participants in blocks of 30s (each image was displayed for 500ms and followed by a 100ms blank screen), and were separated by a 20s fixation block. Each block was repeated twice in each run that lasted for 310s. In the first run of the localizer scan, participants were asked to passively view the images while keeping their fixation at the center of the screen. In the second run of the localizer scan, the first five participants just passively watched the stimuli as they did in the first run, while all the other participants needed to detect two or three animal images that were randomly put into the stream of cars, scrambled cars, and face images by pressing a button with their right hand to ensure participants were paying attention to the stimuli. Face and animal images were purchased from http://www.hemera.com and post-processed using programs written in MATLAB. Car images were picked from the morph space of four prototype cars, and were different from the images used in main experiments. Scrambled car images were generated by scrambling the car images with a grid of 5 by 5 pixel elements. The final size of all images was scaled to 200 by 200 pixels. The stimuli in all scans were presented on a black background using E-Prime (http://www.pstnet.com/products/e-prime/), back-projected on a translucent screen located at the rear of the scanner, and viewed by participants through a mirror mounted on the head coil.

Event-related adaptation Experiments 1 & 2 (displacement detection task)

To probe the effects of categorization training on the tuning of LOC neurons and other brain regions, participants were scanned twice with an fMRI-rapid adaptation (fMRI-RA) paradigm, once prior to training, and again after training. To ensure participants' attention to the stimuli while minimizing task effects that could cause a confounding modulation of fMRI responses (by differentially affecting the experimental conditions of interest), a displacement detection task that was independent of stimulus category membership was adopted: During each trial (except the null trials), two cars were displayed sequentially (300ms each with a 400ms blank screen in-between (Kourtzi and Kanwisher, 2001)) at or close to the center of the screen, followed by a 3000ms blank screen. The second car was presented with a small horizontal displacement relative to the position of the first car, and participants were asked to judge the direction of displacement by pressing a button with their left or right hand depending on the change. For both Experiments 1 and 2, MRI images from six scans were collected. Each run lasted 284s and had two ten second fixation periods, one at the beginning and one at the end. Between the two fixation periods, a total of 66 trials were presented to participants at a rate of one every four seconds. For each run, the data from the first two trials were discarded, and analyses were performed on the data of the other 64 trials – 16 each of the four different conditions defined by the change of shape and category between the two cars: M0 – same category and same shape; $M3_{within}$ – same category and 33.33% shape change; $M3_{between}$ – different category and 33.33% shape change; and null trials (Figure 3A). Trial order was randomized and counterbalanced using M-sequences (Buracas and Boynton, 2002), and number of presentations was equalized for all stimuli in each experiment.

Event-related adaptation Experiment 3 (categorization task)

To assess the neural mechanisms underlying categorization, participants also participated in one more fMRI-RA experiment following Experiment 2, in which two cars were displayed sequentially (300ms each with a 400ms blank screen in-between) at the center of the screen, followed by a 3000ms blank screen during each trial. In these scans, participants needed to judge whether the two cars belonged to the same or different categories by pressing one of the two buttons held in their left and right hand. No feedback was provided to participants. The relationship between the yes/no answers and left/right hand responses was counter-balanced across participants. MRI images from four scans were collected. Each scan lasted 628s and had two 10s fixation periods, one at the beginning and the other at the end. Between the two fixation periods, a total of 127 trials were presented to participants at a rate of one every four seconds. For each run, the data from the first two trials were discarded, and analyses were performed on the data of the other 125 trials – 25 each of the five different conditions defined by the change of shape and category between the two cars: M0 – same category and same shape; $M3_{within}$ – same category and 33.33% shape change; $M3_{between}$ – different category and 33.33% shape change; M6 – different category and 67% shape change; and null trials (Figure 5A). Trial order was randomized and counterbalanced using M-sequences (Buracas and Boynton, 2002).

fMRI acquisition

All fMRI data were acquired at Georgetown University's Center for Functional and Molecular Imaging using an echo-planar imaging (EPI) sequence on a 3.0 Tesla Siemens Trio scanner with a single-channel head coil (flip angle = 90°, TR = 2s, TE = 30ms, FOV = 205, 64×64 matrix). For both functional localizer scans and ER runs, forty-four interleaved axial slices (thickness = 3.2 mm, no gap; in-plane resolution = 3.2×3.2 mm^2) were acquired. At the end, three dimensional T1-weighted MPRAGE images (resolution 1×1×1 mm^3) were acquired for each subject.

fMRI data analysis

All pre-processing and most statistical analyses were done using the software package SPM2 (http://www.fil.ion.ucl.ac.uk/spm/software/spm2/) and its toolboxes. Basically, after discarding the images acquired during the first ten seconds of each run, the images were temporally corrected to the middle slice, then were spatially realigned, unwarped, resliced to 2×2×$2mm^3$, and normalized to a standard MNI reference brain in Talairach space. At the end, two sets of images were created: one set of images were used for the whole-brain analysis and were smoothed with an isotropic 8mm Gaussian kernel, the other set of images were used for the ROI-based analyses and were not smoothed.

The car-selective regions in the LO and face-selective regions in the fusiform area were identified for each individual subject independently with the data from the localizer scans (Grill-Spector et al., 1999;Kourtzi and Kanwisher, 2001). We first modeled the hemodynamic activity for each condition (car, scrambled car, and face) in the localizer scans with the standard canonical hemodynamic response function, then identified the car-selective LO ROI with a contrast of car versus scrambled cars masked by the contrast of car versus baseline ($p<0.00001$ uncorrected), and the face-selective FFA ROI with the contrast of face versus car and scrambled car masked by the contrast of face versus baseline ($p<0.00001$ uncorrected) (see Figure S3 for the results from a representative subject). In total, the right LO and FFA as well as the left LO were reliably identified in all participants and in all experiments. The left FFA was reliably identified in 15 participants in Experiment 1, 16 in Experiment 2, and 14 in Experiment 3. To obtain comparably-sized LO and FFA ROIs across participants, we defined the LO and FFA ROIs by choosing an approximately equal number of contiguous voxels with a minimum of 20 for the car ROI and 80 for the face ROI (Jiang et al., 2006;Murray and Wojciulik, 2004). For details on the ROI selection, see caption of Figure S3) For Experiment 3, to probe the relationship between face responsiveness and car shape selectivity, we defined two additional ROIs in the fusiform face area: (1) a "core" face ROI– a more strictly defined FFA ROI with stricter threshold, which was about half the size of the above-mentioned, more loosely defined FFA ROI for each individual subject; and (2) a "surround" face ROI– an ROI that should respond more weakly to faces by excluding the smaller "core" face FFA ROI from the initial and bigger FFA ROI. The sizes of the two newly defined ROIs were about same within each individual subject ($p>0.4$, paired t-test). For comparison purposes (see text), we further extracted the activation in early visual cortex, which was defined by the contrast of scrambled car versus baseline with a strict threshold (at least $p<10^{-6}$, and $p<10^{-15}$ for most participants).

For the data analysis of Experiments 1, 2, and 3, we first conducted ROI-based analyses using the above-mentioned independently defined ROIs. We extracted the hemodynamic response for each subject in the ROIs using a finite impulse response (FIR) model with the MarsBar toolbox (Brett et al., 2002) and in-house software written in Matlab, and then conducted statistical analyses (repeated measures ANOVA with Greenhouse-Geisser correction, followed by planned t-tests, a=0.05, two-tailed) on the peak values, which were either the values of the 3rd scan or the mean of the 3rd and 4th scan depending on whether the peak lasted for more than one TR.

For the whole-brain analyses on data from Experiments 1, 2, and 3, we modeled fMRI responses with a design matrix comprising the onset of pre-defined non-null trial types (M0, $M3_{within}$, and $M3_{between}$ for Experiments 1 and 2, M0, $M3_{within}$, $M3_{between}$, and M6 for Experiment 3) and movement parameters as regressors using a standard canonical hemodynamic response function (HRF). The parameter estimates of the HRF for each regressor were calculated for each voxel, and then the contrasts at the single subject level were computed and entered into a second-level model in SPM2 (participants as random effects) with additional smoothing (4mm).

For all whole-brain analyses, a threshold of $p<0.001$ (uncorrected) with at least twenty contiguous voxels was used unless otherwise mentioned.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Christian Shelton for the morphing software, Zoe Kourtzi for advice on the fMRI design, and Zoe Kourtzi, Tomaso Poggio and Thomas Serre for comments on earlier versions of this manuscript. This research was supported in part by an NSF CAREER Award (#0449743) and NIMH grants R01MH076281 and P20MH66239. Further support was provided by NCRR grants M01 RR020359 and P30 HD40677.

References

Aizenstein HJ, MacDonald AW, Stenger VA, Nebes RD, Larson JK, Ursu S, Carter CS. Complementary category learning systems identified using event-related functional MRI. J Cogn Neurosci 2000;12:977–987. [PubMed: 11177418]

Ashby FG, Spiering BJ. The neurobiology of category learning. Behav Cogn Neurosci Rev 2004;3:101–113. [PubMed: 15537987]

Brett, M.; Anton, JL.; Valabregue, R.; Poline, JB. Region of interest analysis using an SPM toolbox. NeuroImage; Presented at the 8th International Conferance on Functional Mapping of the Human Brain; Sendai, Japan. 2002. abstract

Buracas GT, Boynton GM. Efficient design of event-related fMRI experiments using M-sequences. Neuroimage 2002;16:801–813. [PubMed: 12169264]

Fang F, Murray SO, He S. Duration-Dependent fMRI Adaptation and Distributed Viewer-Centered Face Representation in Human Visual Cortex. Cereb Cortex. 200610.1093/cercor/bhl053

Freedman DJ, Assad JA. Experience-dependent representation of visual categories in parietal cortex. Nature 2006;443:85–88. [PubMed: 16936716]

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J Neurosci 2003;23:5235–5246. [PubMed: 12832548]

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. Cereb Cortex 2006;16:1631–1644. [PubMed: 16400159]

Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat Neurosci 1999;2:568–573. [PubMed: 10448223]

Gerlach C, Law I, Gade A, Paulson OB. Perceptual differentiation and category effects in normal object recognition: a PET study. Brain 1999;122(Pt 11):2159–2170. [PubMed: 10545400]

Gilaie-Dotan S, Malach R. Sub-exemplar shape tuning in human face-related areas. Cereb Cortex 2007;17:325–338. [PubMed: 16525131]

Grady CL, Horwitz B, Pietrini P, Mentis MJ, Ungerleider LG, Rapoport SI, Haxby JV. Effects of task difficulty on cerebral blood flow during perceptual matching of faces. Hum Brain Mapp 1996;4:227–239.

Grill-Spector K. The neural basis of object perception. Curr Opin Neurobiol 2003;13:159–166. [PubMed: 12744968]

Grill-Spector K, Henson R, Martin A. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci 2006;10:14–23. [PubMed: 16321563]

Grill-Spector K, Knouf N, Kanwisher N. The fusiform face area subserves face perception, not generic within-category identification. Nat Neurosci 2004;7:555–562. [PubMed: 15077112]

Grill-Spector K, Kourtzi Z, Kanwisher N. The lateral occipital complex and its role in object recognition. Vision Res 2001;41:1409–1422. [PubMed: 11322983]

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R. Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 1999;24:187–203. [PubMed: 10677037]

Grill-Spector K, Malach R. The human visual cortex. Annu Rev Neurosci 2004;27:649–677. [PubMed: 15217346]

Jiang X, Rosen E, Zeffiro T, Vanmeter J, Blanz V, Riesenhuber M. Evaluation of a shape-based model of human face discrimination using FMRI and behavioral techniques. Neuron 2006;50:159–172. [PubMed: 16600863]

Kourtzi Z, Betts LR, Sarkheil P, Welchman AE. Distributed neural plasticity for shape learning in the human visual cortex. PLoS Biol 2005;3:e204. [PubMed: 15934786]

Kourtzi Z, Kanwisher N. Representation of perceived object shape by the human lateral occipital complex. Science 2001;293:1506–1509. [PubMed: 11520991]

Kourtzi Z, Tolias AS, Altmann CF, Augath M, Logothetis NK. Integration of local features into global shapes: monkey and human FMRI studies. Neuron 2003;37:333–346. [PubMed: 12546827]

Little DM, Shin SS, Sisco SM, Thulborn KR. Event-related fMRI of category learning: Differences in classification and feedback networks. Brain Cogn 2006;60:244–252. [PubMed: 16426719]

Little DM, Thulborn KR. Correlations of cortical activation and behavior during the application of newly learned categories. Brain Res Cogn Brain Res 2005;25:33–47. [PubMed: 15936179]

Lueschow A, Miller EK, Desimone R. Inferior temporal mechanisms for invariant object recognition. Cereb Cortex 1994;4:523–531. [PubMed: 7833653]

Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annu Rev Neurosci 2001;24:167–202. [PubMed: 11283309]

Miller EK, Li L, Desimone R. Activity of neurons in anterior inferior temporal cortex during a short-term memory task. J Neurosci 1993;13:1460–1478. [PubMed: 8463829]

Miyashita Y, Morita M, Naya Y, Yoshida M, Tomita H. Backward signal from medial temporal lobe in neural circuit reorganization of primate inferotemporal cortex. C R Acad Sci III 1998;321:185–192. [PubMed: 9759339]

Moore CD, Cohen MX, Ranganath C. Neural mechanisms of expert skills in visual working memory. J Neurosci 2006;26:11187–11196. [PubMed: 17065458]

Murray SO, Wojciulik E. Attention increases neural selectivity in the human lateral occipital complex. Nat Neurosci 2004;7:70–74. [PubMed: 14647291]

Nosofsky RM. Attention, similarity, and the identification-categorization relationship. J Exp Psychol Gen 1986;115:39–61. [PubMed: 2937873]

Op de Beeck H, Wagemans J, Vogels R. Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. Nat Neurosci 2001;4:1244–1252. [PubMed: 11713468]

Op de Beeck HP, Baker CI, DiCarlo JJ, Kanwisher NG. Discrimination training alters object representations in human extrastriate cortex. J Neurosci 2006;26:13025–13036. [PubMed: 17167092]

Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 1997;10:437–442. [PubMed: 9176953]

Pollmann S, Maertens M. Shift of activity from attention to motor-related brain areas during visual learning. Nat Neurosci 2005;8:1494–1496. [PubMed: 16205718]

Rainer G, Miller EK. Effects of visual experience on the representation of objects in the prefrontal cortex. Neuron 2000;27:179–189. [PubMed: 10939341]

Reber PJ, Stark CE, Squire LR. Cortical areas supporting category learning identified using functional MRI. Proc Natl Acad Sci U S A 1998;95:747–750. [PubMed: 9435264]

Rhodes G, Byatt G, Michie PT, Puce A. Is the fusiform face area specialized for faces, individuation, or expert individuation? J Cogn Neurosci 2004;16:189–203. [PubMed: 15068591]

Riesenhuber M, Poggio T. Models of object recognition. Nat Neurosci 2000;3(Suppl):1199–1204. [PubMed: 11127838]

Riesenhuber M, Poggio T. Neural mechanisms of object recognition. Curr Opin Neurobiol 2002;12:162–168. [PubMed: 12015232]

Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T. Robust object recognition with cortex-like mechanisms. IEEE Trans Pattern Anal Mach Intell 2007;29:411–426. [PubMed: 17224612]

Shelton C. Morphable Surface Models. Int J Comp Vis 2000;38:75–91.

Sigala N. Visual categorization and the inferior temporal cortex. Behav Brain Res 2004;149:1–7. [PubMed: 14739004]

Sigala N, Logothetis NK. Visual categorization shapes feature selectivity in the primate temporal cortex. Nature 2002;415:318–320. [PubMed: 11797008]

Sunaert S, Van Hecke P, Marchal G, Orban GA. Attention to speed of motion, speed discrimination, and task difficulty: an fMRI study. Neuroimage 2000;11:612–623. [PubMed: 10860790]

Tarr MJ, Gauthier I. FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. Nat Neurosci 2000;3:764–769. [PubMed: 10903568]

Thomas E, Van Hulle MM, Vogels R. Encoding of categories by noncategory-specific neurons in the inferior temporal cortex. J Cogn Neurosci 2001;13:190–200. [PubMed: 11244545]

Vogels R, Sary G, Dupont P, Orban GA. Human brain regions involved in visual categorization. Neuroimage 2002;16:401–414. [PubMed: 12030825]

Yue X, Tjan BS, Biederman I. What makes faces special? Vision Res 2006;46:3802–3811. [PubMed: 16938328]

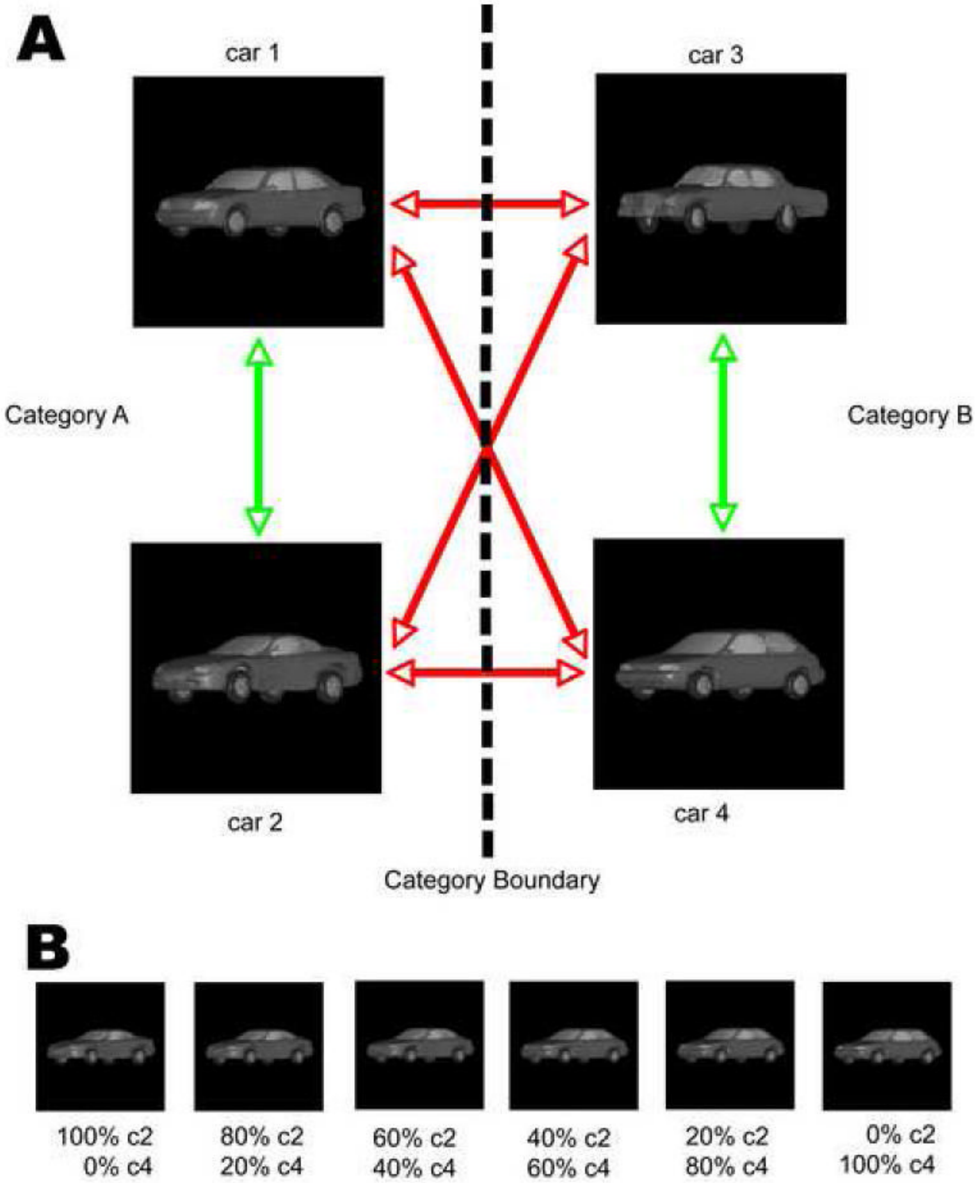

Figure 1. Visual stimuli

(A) Participants learned to categorize randomly generated morphs from the vast number of possible blends of four prototypes. The placement of the prototypes in this diagram does not reflect their similarity. Black lines show cross-category morph lines, gray lines show within-category morph lines. (B) An example morph line between the car2 and car4 prototypes.

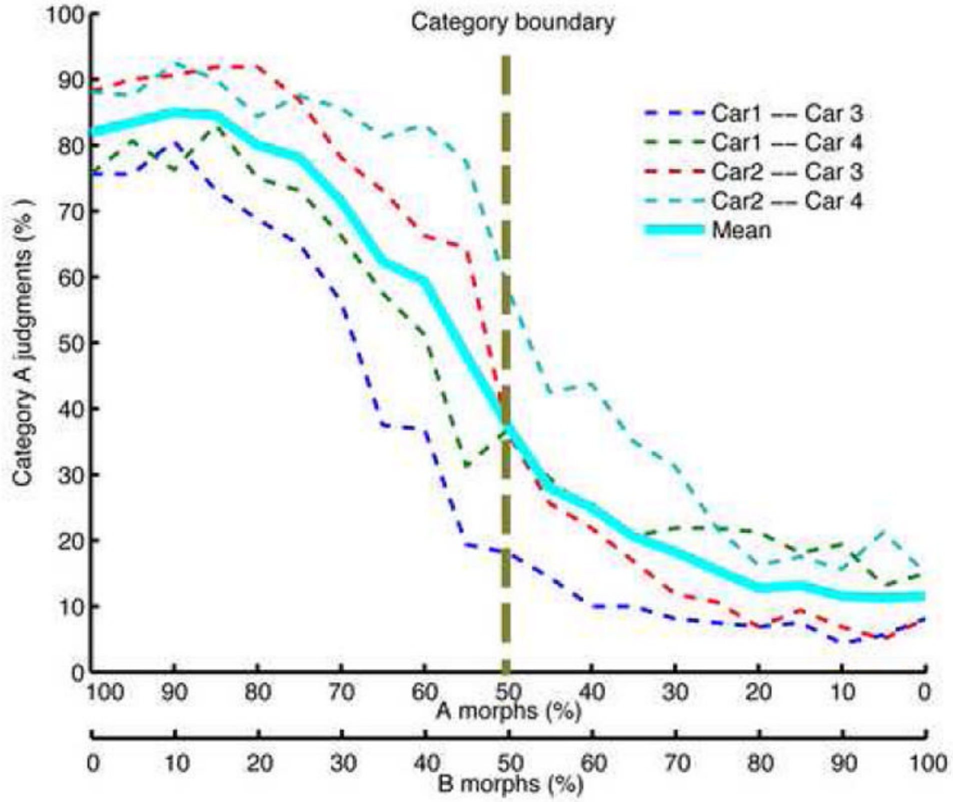

Figure 2. Behavioral categorization data

The average performance (in percent correct on the 2AFC categorization task) along the four cross-category morph lines (dashed), along with the grand average over all morph lines (solid line). The X-axis shows percent morph. To better capture the steep transition around the category boundary that was blurred by averaging across participants and morph lines, we also fitted sigmoid functions to individual subject performances and then averaged across the fitted sigmoid parameters, see Figure S2.

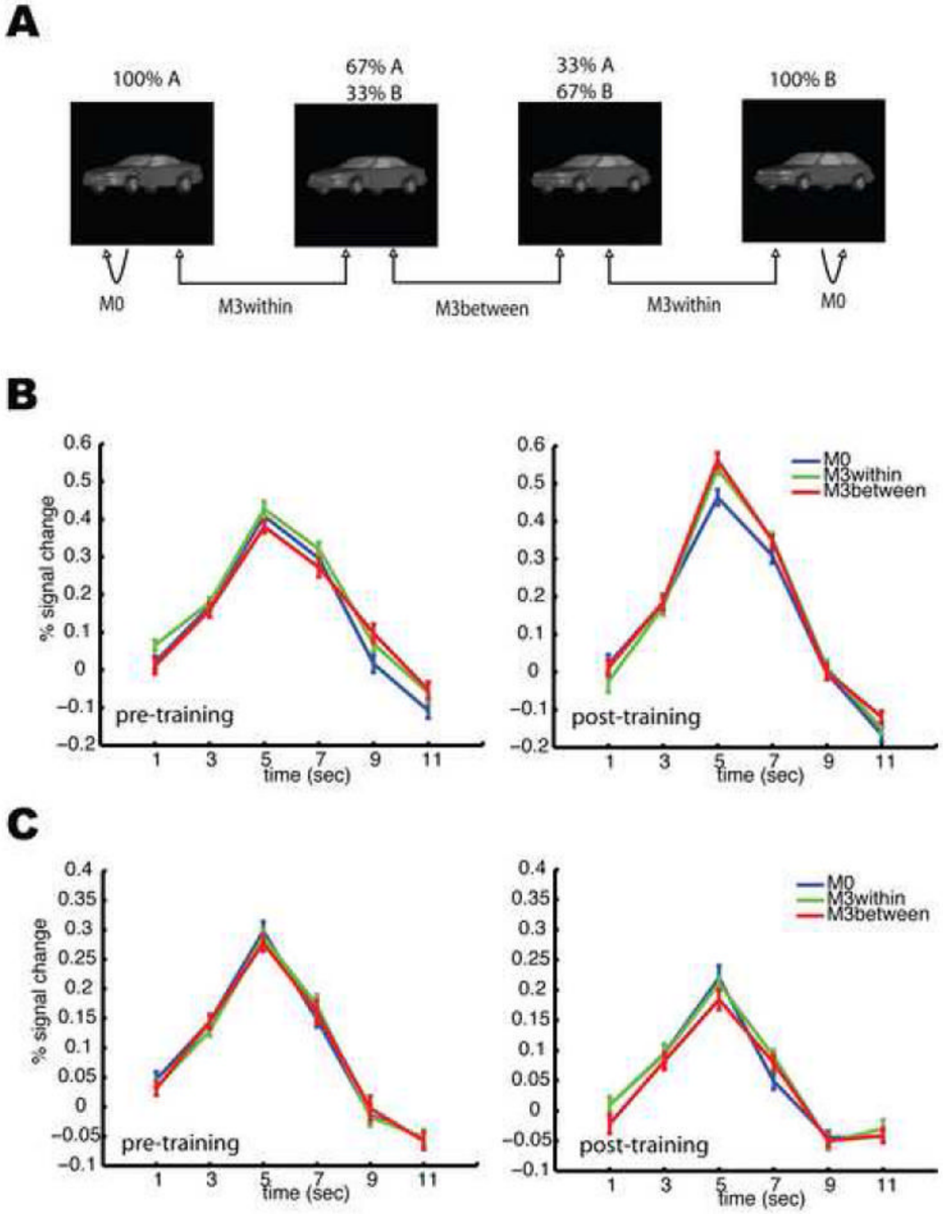

Figure 3. fMRI-RA Experiments 1 (pre-training) and 2 (post-training) in which participants performed a displacement judgment task

Three conditions, M0, M3$_{within}$, and M3$_{between}$, were tested. Using one morph line as an example, (A) shows how stimulus pairs were constructed. (B) shows the mean fMRI response in the rLO pre- (left) and post-training (right). (C) shows the mean fMRI responses in the rFFA pre- (left) and post-training (right). A significant difference of peak values among the three conditions was only observed in the rLO after training. Error bars show within-subject SEM.

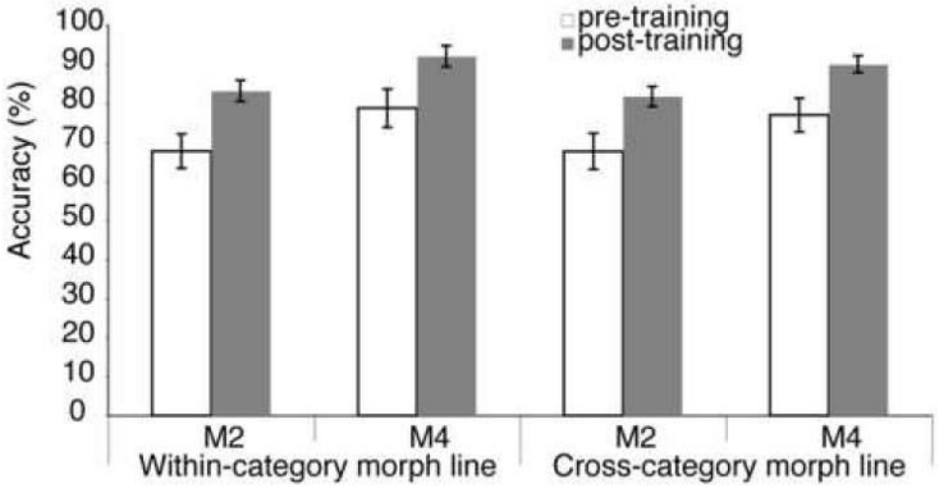

Figure 4. Psychophysical performance on the car discrimination task

Participants (n=13, see Methods) were tested on a 2AFC discrimination task using pairs of car stimuli chosen from all six morph lines, including two within-category morph lines and four cross-category morph lines (see Figure 1). Testing was done both before ("pre-training") and after ("post-training") categorization training. Match and non-match stimuli in each trial could either differ by 20% (M2) or 40% shape change (M4). An ANOVA with training (pre- versus post-training), morph lines (within- versus cross-category morph lines), and morph step difference between match and non-match choice stimuli (M2 vs. M4) as repeated measures revealed significant effects of category training, F(1,12)=7.358,p=0.019, and morph step difference, F(1,12)=172.129, p<0.001, but not for morph line, F(1,12)=2.633, p=0.131. Importantly, there were no significant interactions, in particular not for training effect vs. morph line, demonstrating that category learning improved discrimination of stimuli in general and not just for the category-relevant morph lines. Error bars show SEM.

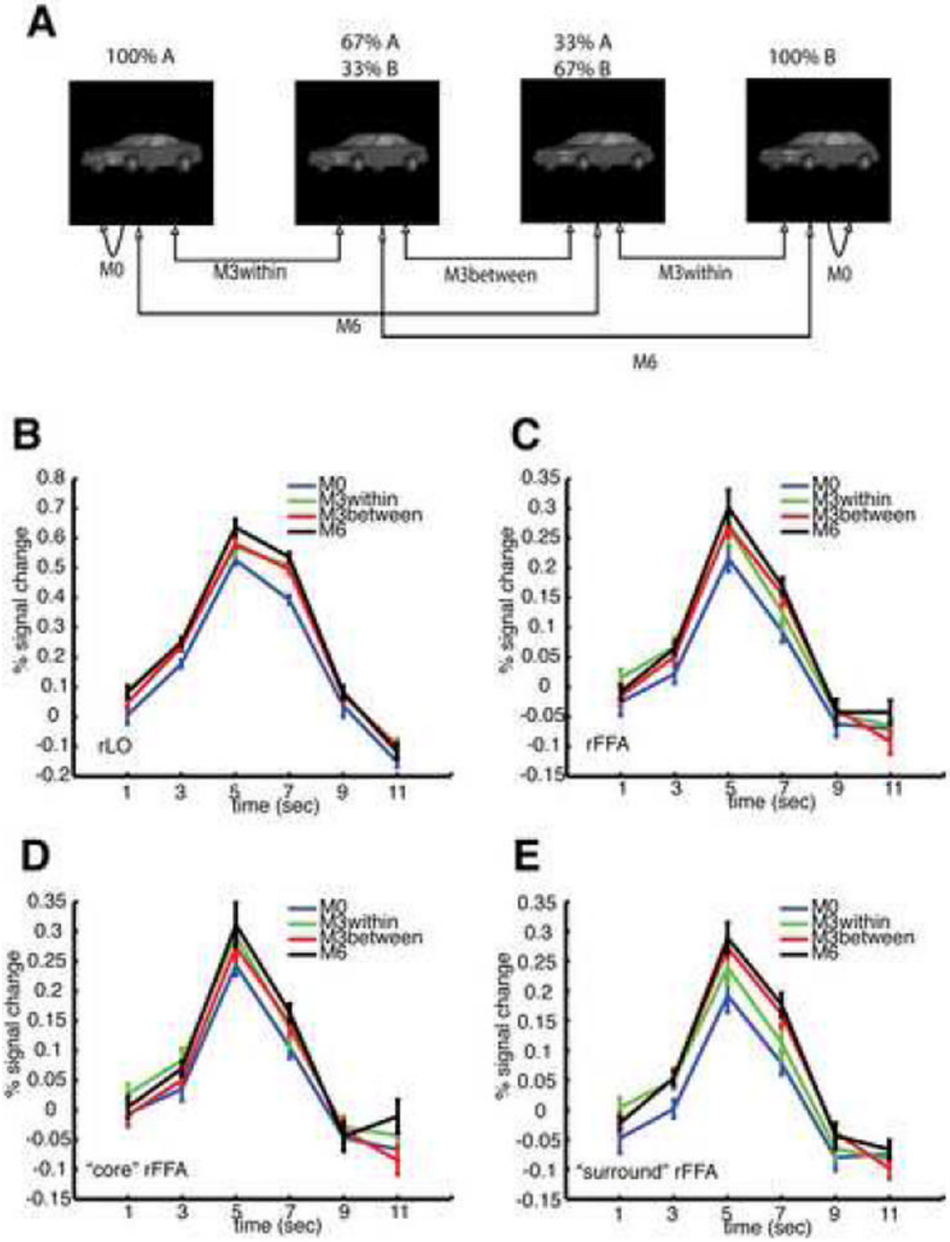

Figure 5. fMRI-RA Experiment 3, in which participants needed to perform a same/different categorization task on the pair of stimuli in each trial

Four conditions (M0, $M3_{within}$, $M3_{between}$, and M6) were tested. The choice of stimuli for each condition is shown in (A). A significant difference of peak values was found in rLO (B), but not in rFFA (C), nor in the right “core FFA” (D), though a marginal effect was found in the right “surround FFA” (E) (see text). The legends for (B)-(E) are the same and shown in (E). Error bars show within-subject SEM.

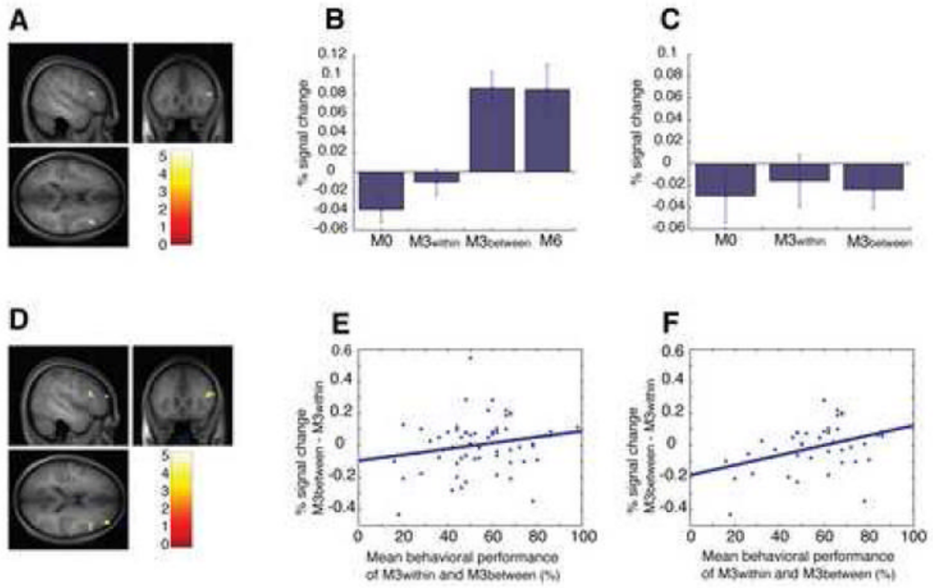

Figure 6. Activation in the rLPFC ROI

(A) The rLPFC ROI defined by the comparison of $M3_{between}$ versus $M3_{within}$ of the morph line on which participants had the best behavioral performance ($p<0.001$, uncorrected, size: $280mm^3$, shown as sagittal, coronal, and axial sections on an average anatomical image generated from the individual T1-weighted images of the 16 participants in Experiment 3, same for (D)), and mean signal change for trials along this morph line at this ROI for the conditions of M0, $M3_{within}$, $M3_{between}$, nd M6 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 $M3_{between}$) as repeated measures found significant differences for the data set of Experiment 3 (B), $p<0.00001$, but not for the data set of Experiment 2 (C) $p>0.5$. Similar activation patterns were also found when the rLPFC ROI was defined by the comparison of M6 versus M0, and M6 and $M3_{between}$ versus $M3_{within}$ and M0 of same morph line for each individual subject (Figure S14). (D) The rLPFC ROI defined by the comparison of M6 versus M0 of the morph line on which participants had the best performance ($p<0.001$, uncorrected, size: $512mm^3$). We then calculated the BOLD contrast response difference between the $M3_{within}$ and $M3_{between}$ conditions for each morph line and subject (y-axis), and plotted this index against the mean behavioral accuracy on these conditions inside the scanner (x-axis). (E) shows the data for all participants (N=16) and the regression line ($r=0.206$, $p=0.102$). (F) shows the data for the subgroup of participants (N=11) with above-threshold behavioral performance on the M0 and M6 conditions (see text) along with the regression line ($r=0.409$, $p<0.01$). Error bars show within-subject SEM.

Table 1

Brain regions showing stronger activations to pairs of cars belonging to different categories than to pairs belonging to the same category

Region	mm^3	Z^{max}	MNI Coordinates X	 Y	 Z
M6 > M0					
R Inf Temporal	1608	4.69	50	-48	-20
R Inf/Mid Frontal Gyrus/Insula	**8512**	**4.65**	**42**	**34**	**16**
		4.07	**38**	**18**	**14**
		3.87	**54**	**36**	**16**
R Mid Cingulum	264	4.55	18	-24	36
L Inf/Sup Parietal	12688	4.49	-32	-46	38
		4.46	-24	-52	38
		4.42	-28	-52	46
R Sup Occipital/R Inf/Sup Parietal	9400	4.31	44	-40	46
		4.15	34	-44	36
		4.08	30	-74	34
L Cerebum	504	4.14	-8	-22	-44
L/R Sup Motor Area	1240	4.03	6	12	52
L Precentral	560	3.99	-42	4	34
R Inf Temporal	400	3.79	44	-30	-22
L Inf Temporal	552	3.63	-46	-54	-22
		3.42	-40	-58	-30
R Sup Frontal	256	3.57	28	8	68
R Mid Frontal	392	3.51	50	52	14
		3.30	46	58	22
R Inf/Mid Frontal	504	3.48	36	6	34
R Inf Frontal	168	3.45	38	0	22
R Mid Frontal	168	3.44	42	4	58
R Brainstem	168	3.37	8	-24	-32
L/R Sup Motor Area	192	3.26	2	-2	62
L Inf Frontal Gyrus*	136	3.37	-48	20	16
R Inf/Mid Occipital*	136	3.40	40	-84	-2
R Inf Occipital/Temporal*	152	3.33	58	-64	-18
M6 & M3$_{between}$ > M3$_{within}$ & M0					
R Inf/Mid Frontal Gyrus	**1848**	**4.09**	**48**	**32**	**16**
		3.56	**52**	**30**	**26**
		3.55	**46**	**24**	**22**
R Inf Temporal	312	3.75	48	-52	-20
L Inf Parietal	936	3.71	-40	-40	42
R Insula	280	3.40	30	26	8
R Mid Cingulum	264	3.34	6	24	38
L Inf Frontal Gyrus*	96	3.31	-52	20	24
M0 > M6					
L Angular	384	3.64	-52	-70	32
L/R Med Frontal	712	3.64	0	34	-8
M3$_{within}$ & M0 > M6 & M3$_{between}$					
L Sup Occipital	216	3.71	-18	104	18

[*] cluster size is smaller than 20 but larger than 10

Table 2

Brain regions showing stronger activation to pairs of cars belonging to different categories than to pairs belonging to same category, even when the intra-pair shape change was the same ($M3_{between}$ versus $M3_{within}$ trials of the morph line on which participants had the best performance, see text)

Region	mm^3	Z_{max}	MNI coordinates X	 Y	 Z
M6 > M0					
R Mid Occipital	440	3.84	30	-74	36
R Mid Frontal	416	3.68	42	52	16
L Inf Parietal	264	3.65	-50	-34	44
R Inf Frontal Gyrus	**512**	**3.45**	**46**	**24**	**20**
M6 & $M3_{between}$ > $M3_{within}$ & M0					
R Inf Frontal Gyrus	**1088**	**4.18**	**44**	**28**	**12**
R Mid Occipital	232	3.52	34	-68	34
$M3_{between}$ > $M3_{within}$					
R Inf Frontal Gyrus	**280**	**3.87**	**48**	**26**	**16**