

## MIT Open Access Articles

### *Recognition alters the spatial pattern of fMRI activation in early retinotopic cortex*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Hsieh, P.-J., E. Vul, and N. Kanwisher. "Recognition Alters the Spatial Pattern of fMRI Activation in Early Retinotopic Cortex." *Journal of Neurophysiology* 103.3 (2010): 1501–1507. Web.

**As Published:** <http://dx.doi.org/10.1152/jn.00812.2009>

**Publisher:** American Physiological Society

**Persistent URL:** <http://hdl.handle.net/1721.1/72170>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike 3.0



**Recognition alters the spatial pattern of fMRI activation in early retinotopic cortex**

**Po-Jang Hsieh\*, Ed Vul, and Nancy Kanwisher**

Department of Brain and Cognitive Sciences, McGovern Institute, MIT,  
Cambridge, MA 02139, USA.

Running title: Hierarchical Inference in the Visual System

Keywords: hierarchical inference; feedback; perceptual learning; top-down modulation

\*To whom correspondence should be addressed.  
[pjh@mit.edu](mailto:pjh@mit.edu)

Early retinotopic cortex has traditionally been viewed as containing a veridical representation of the low-level properties of the image, not imbued by high-level interpretation and meaning. Yet several recent results indicate that neural representations in early retinotopic cortex reflect not just the sensory properties of the image, but the perceived size and brightness of image regions. Here we used fMRI pattern analyses to ask whether the representation of an object in early retinotopic cortex changes when the object is recognized compared to when the same stimulus is presented but not recognized. Our data confirmed this hypothesis: the pattern of response in early retinotopic visual cortex to a two-tone ‘Mooney’ image of an object was more similar to the response to the full grey scale photo version of the same image when observers knew what the two-tone image represented than when they did not. Further, in a second experiment, high-level interpretations actually overrode bottom-up stimulus information, such that the pattern of response in early retinotopic cortex to an identified two-tone image was more similar to the response to the photograph version of that stimulus than it was to the response to the identical two-tone image when it was not identified. Our findings are consistent with prior results indicating that perceived size and brightness affect representations in early retinotopic visual cortex, and further show that even higher-level information--knowledge of object identity – also affects the representation of an object in early retinotopic cortex.

## Introduction

Early retinotopic cortex (i.e. V1/V2/V3) has traditionally been viewed as an approximately veridical representation of the low-level properties of the image, which provides the input to higher cortical areas where meaning is assigned to the image. But how distinct are data and theory in visual cortex? Recent evidence has shown that activity in early retinotopic cortex can reflect mid-level visual computations such as contour completion (Stanley and Rubin, 2003) and figure-ground discrimination (Hupé et al., 1998; Heinen, et al., 2005; Huang and Paradiso, 2008), and even perceived size (Murray et al., 2006) and lightness constancy (Boyaci et al., 2007). Further, at least some of the responses in early retinotopic cortex are likely affected by feedback from higher areas, including attentional modulations (Datta and DeYoe, 2009; Fischer and Whitney, 2009; Ress et al., 2000), stimulus reward history information (Serences, 2008), and lower responses when a shape is perceived compared to when it is not (Murray et al., 2002; Fang, Kersten, & Murray, 2008a). Here we ask whether retinotopic cortex contains information about the perceived identity of an object. We unconfound the perceived identity of the object from its low-level visual properties by comparing the response to a Mooney image before and after it is disambiguated.

Specifically, we scanned subjects while they viewed Mooney images they could not interpret, then greyscale versions of those same images, then the original Mooney images which subjects could now understand. We found that the spatial pattern of activation in early retinotopic cortex of a Mooney image more closely resembles the activation pattern from the unambiguous greyscale version of that same image when the Mooney image is understood than when it is not. This finding goes beyond prior findings indicating that the representation of a stimulus in early retinotopic cortex reflects higher-level information such as perceived size (Murray et al., 2002; Fang et al, 2008a), brightness (Boyaci et al., 2007), and grouping (Fang et al, 2008b; Murray et al, 2002, 2004) by showing that the activation pattern of a stimulus in early retinotopic cortex also changes when it is recognized compared to when it is not. Further, our use of pattern analysis enables us to show that (1) the specific way the activation pattern (i.e., representation) in early

retinotopic cortex changes is to become more like the unambiguous stimulus, and (2) these higher-level effects on pattern information are dissociable from the mean changes that have been reported previously.

## **Methods**

**Participants.** Seven subjects (age range: 20-30) were run in the blocked design experiment (Experiment 1), and eight subjects (age range: 20-30) were run in the event-related design experiment (Experiment 2). All subjects had normal or corrected-to-normal vision. All gave written consent within a protocol passed by the Massachusetts Institute of Technology committee on the Use of Humans as Experimental Subjects. Subjects were paid sixty dollars per session.

**Experimental procedures.** In Experiment 1 (**Figure 1a**), each subject participated in 12 runs. In the first 4 runs (phase1: the “Mooney1” condition), subjects were presented with 2 ambiguous two-tone images containing unidentifiable objects (7x7 in visual degrees). In the following 4 runs (phase2: “Photo”), subjects were presented with greyscale versions of the same stimuli in which the object could be clearly identified. In the last 4 runs (phase3: “Mooney2”), the original two-tone images were presented again, but now the pictured object could be easily recognized due to the experience of the greyscale images. There were a total of 8 blocks of stimulus presentation in each run. During each block (20 seconds), the same image (a fish or a camel) was presented at 1 Hz at the center of the screen (remaining on for 500ms, followed by a blank inter-stimulus interval of 500ms). The 2 kinds of block (fish-block and camel-block) were interleaved with blank blocks (20 seconds each); order of blocks was counterbalanced across runs. Each picture had two possible contrasts (high or low). For each stimulus presentation, participants were required to indicate via a 2-button response box whether the stimulus had a relatively high or low contrast.

In Experiment 2, all the procedures were identical to those in Experiment 1 except that there were a total of 60 events of stimulus presentation in each run (30 for each stimulus condition). During each stimulus presentation, one of the two images was presented at the center of the screen for 180ms. The order of the trials was optimized

using an optimal sequencing program Optseq2 (NMR Center, Massachusetts General Hospital, MA, USA). Each run lasted 5 min 15 s. More than one-third of the scanning time (3/7) consisted of null events that were randomly inserted between trials. For each subject, 2 images were randomly selected from a set of 8 images (**Figure 1b**). Before the experiment, subjects were shown briefly with the two randomly selected pictures (remembered by subjects as pictures 1 and 2) and verbally confirmed that they could not recognize the images. After subjects finished phase1 and before they started phase2, we presented the photo version of the images during the break to make sure they can recognize the objects in the photo images. After subjects finished phase2 and before started phase3, we present the photo version of the images and the Mooney images side by side during the second break to make sure they could recognize that the objects in the Mooney images are the same as in the Photo images. Subjects were also asked after phase3 if they could identify the stimuli during phase3. For each stimulus presentation during the experiment, participants were required to indicate which one of the two images (pictures 1 or 2) was presented via a 2-button response box.

### ***ROI identification***

The localizer scans were run as described previously (Kourtzi and Kanwisher, 2001). Functional localization of the ROIs was based on independent runs (minimum=1; maximum=3) of four 20-second blocks of grayscale faces, scenes, common objects and scrambled objects (4 blocks per category per run). The critical region of interest (ROIs) identified in the independent localizer scan was the LOC (defined as the region that responded more strongly to images of intact objects than to images of scrambled objects (Malach et al., 1995). The foveal confluence (Dougherty et al., 2003) was defined as a small region at the posterior end of the calcarine sulcus, functionally constrained to the regions that responded more strongly to any images of (intact or scrambled) than to fixation (**Figure 2**). Specifically, we overlaid each subject's functional contrast map (intact objects > scrambled objects) on top of his/her inflated brain and then selected voxels located (1) around the posterior tip of the calcarine sulcus and (2) within the functional contrast map. Since our stimuli extended about 3.5 degrees from the fovea and most of the features in the image were located in the center, we decided to be

conservative and restricted the size of the foveal confluence so that it is smaller than the minimum size of foveal confluence reported in Dougherty et al. (2003). An additional annulus region surrounding the foveal confluence was identified as a control ROI. Since the annulus ROI was chosen to be a control region, the size of this ROI was selected to be larger than the foveal confluence (to gain power for the correlation analysis). The average size of the foveal confluence ROI in Experiments 1 and 2 were 161 ( $\pm 19$  s.e.m.; range: 103-257) and 153 ( $\pm 15$ ; range: 103-228) voxels. Pictures for individual ROIs are provided in **Supplementary Figure 1**.

**Data Analysis.** Scanning was done on the 3T Siemens Trio scanner in the McGovern Institute at MIT in Cambridge (Athina A. Martinos Imaging Center). Functional MRI runs were acquired using a gradient-echo echo-planar sequence (TR=3s, TE=40ms,  $1.5 * 1.5 * 1.5$  mm + 10% spacing). Forty slices were collected with a 32-channel head coil. Slices were oriented roughly perpendicular to the calcarine sulcus and covered most of the occipital and posterior temporal lobes, as well as some of the parietal lobes.

fMRI data analysis was conducted using freesurfer (<http://surfer.nmr.mgh.harvard.edu/>). The processing steps for both the localizer and experimental runs included motion correction and linear trend removal. The processing for the localizer also included spatial smoothing with a 6mm kernel. A gamma function with  $\delta = 2.25$  and  $\tau = 1.25$  was used to estimate the hemodynamic response for each condition in the localizer scans.

Correlation analysis was conducted on the beta weight for each condition in each voxel with a standard multivariate pattern analysis method (Haxby et al., 2001). The mean response in each voxel across all conditions was subtracted from the response to each individual condition for each half of the data before calculating the correlations. Spatial patterns were extracted from each set of data (Mooney1, Photo, and Mooney2) separately for each combination of ROI and stimulus type (e.g., fish versus camel). Within each ROI we then computed the correlation between the spatial patterns of Mooney1 and Photo runs from the same stimulus category (for example, between Mooney1-fish and Photo-fish, and between Mooney1-camel and Photo-camel). The same correlation was computed between the spatial patterns of Mooney2 and Photo runs, and

between the spatial patterns of Mooney2 and Mooney1 runs. These correlations were computed for each subject and then averaged across subjects by conditions. Note that our method is equivalent to that used by Haxby et al. (2001), in which a given region of interest is deemed to contain information about a given stimulus discrimination (Stimulus A versus stimulus B) if the pattern of response across voxels in that region is more similar for two response patterns when they were produced by the same stimulus than when they were produced by two different stimuli, i.e. if  $\text{mean}(r(A1,A2), r(B1,B2)) > \text{mean}(r(A1,B2), r(A2,B1))$ . In our method, we directly computed the correlation between the spatial patterns of Mooney and Photo runs from the same stimulus category (for example, between Mooney-fish and Photo-fish, and between Mooney-camel and Photo-camel). Although, conceptually, saying that an ROI contains discriminative information about stimulus A versus stimulus B (as in Haxby et al., 2001), is very different from saying that the activation pattern in this ROI is more similar between Phase 2 and Phase 3 than between Phase 1 and Phase 2 as in our method, mathematically they amount to the same thing because of the normalization procedure (activation values were normalized to a mean of zero in each voxel across categories by subtracting the mean response across all categories) and the fact that there were only 2 categories.

## Results

We tested whether top-down signals alter low-level representations to match high-level interpretations in an fMRI experiment with three phases (**Figure 1a**). Subjects viewed (i) two different two-tone images that they could not identify (“Mooney1”); (ii) the easily identifiable grey-scale photographic versions of the same images, e.g. a camel and a fish (“Photo”); and (iii) the original two-tone images which could now be easily recognized because of experience with the corresponding photograph (“Mooney 2”). If top-down signals make representations in early retinotopic cortex areas more closely match learned interpretations, we would predict that the fMRI response patterns in early retinotopic cortex will be more similar between the Photo and Mooney2 phases for a given stimulus, than between the Photo and Mooney1 phases for the same stimulus.



To address this question, we used a standard multivariate pattern analysis method (Haxby et al., 2001) (See Methods). This analysis was applied to three (independently defined) regions of interest (ROI): the lateral occipital complex (LOC); the foveal confluence (FC), an early retinotopic area at the occipital pole where the foveal representations of visual areas V1, V2, and V3 converge; and an annulus ROI surrounding the FC (**Figure 2**).

### *Pattern information in LOC*

First we considered ventral visual area LOC, where the fMRI response has been shown previously to change after subjects recognized Mooney images (Dolan et al., 1997). Consistent with prior work, we found that the pattern of response across voxels in the LOC reflected subject's interpretation of object identity in the first, blocked design experiment (see **Figure 1a**). That is, the pattern of response for the camel during the Photo phase was more similar to the camel in the Mooney2 phase than the camel in the Mooney1 phase (**Figure 3a**; paired  $t(6)=3.69$ ,  $p=0.010$ ). This result indicates that the representation of a given Mooney image in LOC is more similar to the representation of the photographic version of the same image after subjects have learned what that Mooney image represents than before. Note that previous studies showed only that perceptual learning of faces or objects enhanced activity in ventral visual areas (Dolan et al., 1997; George et al., 1999; James et al., 2002; McKeef and Tong, 2006; Deshpande et al. 2009). Here we showed for the first time that the activation pattern of a given Mooney image actually changed to become more similar to the activation pattern of the photographic version of the same image.

### *Pattern information in early visual cortex*

Our key prediction is that this pattern of correlations should exist in early retinotopic cortex as well: i.e., the high-level interpretation of a stimulus will modify the representation of that stimulus in early retinotopic cortex to make it more closely match what would be expected given that interpretation. Confirming our predictions, we find the

same effect in early retinotopic cortex: the Mooney2-Photo correlation was greater for corresponding stimuli than the Mooney1-Photo correlation (**Figure 3b**; paired  $t(6)=2.48$ ,  $p=0.048$ ). This result indicates that the representation in early retinotopic cortex for an ambiguous Mooney image is more similar to the early retinotopic representation of a grey-scale version of the same image when subjects interpret the Mooney as containing the information in the photo than when they do not. Analysis of the correlations for non-corresponding stimuli (camel to fish) reveals that our results do not simply reflect overall higher responses between stimuli in Photo-Mooney2 versus Photo-Mooney1 (see Methods). Note that, for both the foveal confluence ROI and LOC, the Mooney2-Mooney1 correlation is the highest of all the comparisons. This is not surprising given that the bottom-up information is identical in Mooney2 and Mooney1.

Our finding suggests that the higher-level interpretation of a stimulus apparently alters the representation of that stimulus in early retinotopic cortex to make it match the representation consistent with the higher-level interpretation. However, before this conclusion can be accepted, we must address an alternative account subjects may have fixated different parts of the stimuli (e.g., the eyes) when they could identify the stimulus (i.e. in the Photo and Mooney 2 phases), compared to when they could not (the Mooney 1 phase), making the patterns of early retinotopic response more similar for the photograph and Mooney 2 cases (which are both recognized) than either is to the Mooney 1 case (which is not recognized). To test this possibility, we conducted a second experiment with an event-related design in which the two different stimuli were interleaved in a random order within each block, and each stimulus was displayed for only 180 ms. Thus, subjects could not predict which stimulus would appear next, and they did not have time to make a saccade while a stimulus was presented. Nonetheless, we replicated the findings from our first experiment: we again found a higher correlation between Photo and Mooney 2 than between Photo and Mooney 1 (**Figure 5**; paired  $t(7)=6.77$ ,  $p<0.001$ ). Thus, our results cannot be explained by differential eye fixation patterns or predictive signals.

With Experiment 2, we further tested whether the degree to which the early retinotopic

cortex representation matches the interpretation (top-down priors) as opposed to the data (bottom-up input) increases as the amount of bottom-up information decreases. We suspected that in the event-related design, when stimuli are presented briefly (180 ms, compared to 500 ms in Experiment 1) and less stimulus information is available, the early retinotopic representation of Mooney2 would be even more similar to the Photo than Mooney1. This possibility is consistent with many models of vision positing that V1 reflects a combination of high-level information and low-level properties of an image. For example, one popular proposal about the computational role of top-down signals in the visual system is that these signals are the brain's implementation of hierarchical Bayesian inference, i.e., they are the priors for the low-level visual areas. On this account, if prior information is particularly strong, and incoming data are particularly ambiguous, then the patterns of response between Mooney2 and the Photo in early retinotopic cortex might be even more similar than between Mooney2 and Mooney1.

Indeed, our data are consistent with this interpretation: presenting less bottom-up information (in Experiment 2 compared to the Experiment 1) results in a proportional increase in the Mooney2-Photo correlation relative to the Mooney2-Mooney1 correlation (chi square sign-test of the interaction between Experiments 1 and 2 and Mooney2-Photo versus Mooney2-Mooney1,  $p=0.02$ ). This interaction between amount of stimulus information and amount of top-down information supports that the hypothesis that the modulation of early retinotopic cortex in this task reflects not just a superposition but an integration of top-down interpretation (priors) with the bottom-up data. However, further work is necessary to test whether bottom-up and top-down inputs to retinotopic cortex interact.

#### *No mean response difference in LOC and foveal confluence ROI*

For the blocked-designed experiment (Experiment 1), the mean responses in both foveal confluence ROI and LOC (mean beta weights averaged within ROI) do not differ between Mooney1 and Mooney2 (paired  $t(6)<1.23$ ,  $p>0.26$ ), indicating that the pattern information is independent of the mean response (**Figure 4**). For the event-related

experiment (Experiment 2), there is a trend of a higher response in FC for Mooney2 than Mooney 1 (**Figure 6**), but it does not reach significance (paired  $t(7)=1.86$ ,  $p=0.11$ ). This result contrasts with previous studies (McKeeff and Tong, 2007), that recognized Mooney images induce larger signals in higher visual areas than unrecognized Mooney images. We suspect that this discrepancy arises from the presentation of only two Mooney images in our experiment, which might lead to stronger adaptation effects than in these prior experiments.

*The pattern of response did not change in the annulus ROI surrounding the foveal confluence*

We checked the anatomical specificity of our effect by analyzing a control ROI defined as an annulus surrounding the foveal confluence ROI. We found none of the critical effects in this annulus ROI: correlations between Photo and Mooney 2 were identical to those between Photo and Mooney 1, for both Experiment 1 (paired  $t(6)=0.07$ ,  $p=0.94$ , **Figure 7a**) and Experiment 2 (paired  $t(7)=0.70$ ,  $p=0.51$ , **Figure 7b**). The two-way ANOVA (region x experiment) revealed a significant region effect ( $F(1,25)=11.84$ ,  $p=0.002$ ). These results suggest that our finding is not resulting from a spillover from higher-level areas.

## **Discussion**

Our results show that the recognition of an object alters its representation in early retinotopic cortex. Specifically, the early retinotopic representation of an ambiguous Mooney stimulus becomes more similar to the representation of an unambiguous grayscale version of the same stimulus when the Mooney stimulus is recognized compared to when it is not. Our finding goes beyond prior reports of correlations between neural responses and subjective percepts (McKeeff and Tong, 2006; Serences and Boynton, 2007), which might reflect trial-to-trial fluctuations in neural noise in early retinotopic visual areas propagating in a feedforward fashion to higher areas, producing different subjective percepts. In contrast, in our study we directly manipulated the high-

level interpretation of the stimulus, and observed its effect in early retinotopic regions, enabling us to infer the causal direction of our effect as reflecting feedback from higher areas (such as LOC) to early retinotopic visual areas (Murray et al., 2002; Williams et al., 2008). Our findings therefore suggest that representations in early retinotopic visual cortex reflect an integration of high-level information about the identity of a viewed object and low-level properties of that object.

Our finding is consistent with other evidence that responses in early visual cortex can reflect influences of mid- and high-level visual processes, including figure-ground segregation (Hupé et al., 1998; Heinen, et al., 2005; Huang and Paradiso, 2008), contour completion (Stanley and Rubin, 2003), identification of object parts (McKeeff, T. J. & Tong, 2006), perceived size (Murray et al., 2006, Fang et al., 2008a), lightness constancy (Boyaci et al., 2007), and grouping (Murray et al., 2002; Fang et al., 2008b). Our finding goes beyond these other findings by showing an effect of object recognition on the representation of a stimulus in early retinotopic cortex. Further, our use of pattern analysis enables us to show that the specific nature of change in early retinotopic cortex is that the representation becomes more like that of the unambiguous (grayscale) stimulus when the stimulus is recognized (Mooney2) compared to when it is not (Mooney1).

Perhaps the closest prior finding to ours is the report that mean responses in early retinotopic cortex are higher for stimuli that have not been grouped or recognized than stimuli that have (Murray 2002, Fang et al 2008a). We did not find this previously-reported difference in mean responses in our study, despite finding that the activation pattern in early retinotopic cortex changes after the subjects recognized the stimuli. This discrepancy between our results and the prior studies may be due to the fact that our subjects were performing an orthogonal task whereas some of the previously-reported changes in early retinotopic cortex due to high-level interpretations are dependent on attention (Fang et al., 2008b). In any case, the fact that we found a change in the pattern response without a change in the mean response indicates that the two phenomena are distinct.

Our findings are consistent with several theoretical frameworks. According to one, the visual system implements hierarchical Bayesian inference (Lee and Mumford, 2003; Rao, 2005), such that each visual area integrates priors (top-down signals) with new data (bottom-up input) to infer the likely causes of the input. According to the framework of perception as Bayesian inference (Knill and Richards, 1996), prior knowledge about the structure of the world provides the additional constraints necessary to solve the problem of vision, that is, to infer the causes in the world of particular retinal images (Lee, 1995; Mumford, 1992). Although numerous proposals have been made about how Bayesian inference could be implemented in the brain (Lee and Mumford, 2003), neural evidence for such computations has proven elusive (Knill and Pouget, 2004). Our findings provide evidence for one specific implementation of Bayesian inference in the visual system: a hierarchical system in which each layer integrates prior information from higher levels of abstractions with data from lower layers of abstraction (Lee and Mumford, 2003).

Other interpretations of our finding include the possibility that the changes in early retinotopic cortex that arise when an image is understood reflect interpretation-contingent visual attention to specific retinotopic coordinates or mental imagery (Klein et al., 2004; Mohr et al., 2009). These effects might occur after recognition is complete, in which case they would constitute an alternative to the Bayesian account. However, attention/imagery accounts, according to which these mechanisms play a role in recognition itself, need not be alternatives to the Bayesian account. For example, attentional allocation to specific features at particular spatial locations based on high-level information is one method of combining top-down priors and bottom-up information, and can be seen as a mechanism by which hierarchical Bayesian inference is accomplished. Indeed, attention to specific spatial coordinates has been postulated as the mechanism driving contour completion (Stanley and Rubin, 2003), identification of object parts (McKeeff and Tong, 2006), and figure-ground segregation (Hupé et al., 1998). These computations all require constraints in addition to the bottom-up input, and can be computationally formalized as hierarchical Bayesian inference. A higher-level interpretation (e.g. of object identity) is a natural source of additional information that can help to segregate figure from ground, fill in surfaces, complete contours, discern object parts, or otherwise sharpen early visual

representations: If we learn that an image represents a face, we can infer where to expect contours, parts, figure, and ground. Moreover, the higher-level interpretation is often easier to determine than the details, both for people and Bayesian models (Kersten et al., 2004). We propose that this “blessing of abstraction” allows the visual system to efficiently parse visual scenes by exploiting its hierarchical structure: higher visual areas can quickly infer the global structure of the scene or object, and can then provide these interpretations as priors to the lower visual areas, thus disambiguating otherwise ambiguous input data (Bar, 2003, 2006; Hochstein and Ahissar, 2002; Ullman, 1995).

## REFERENCES

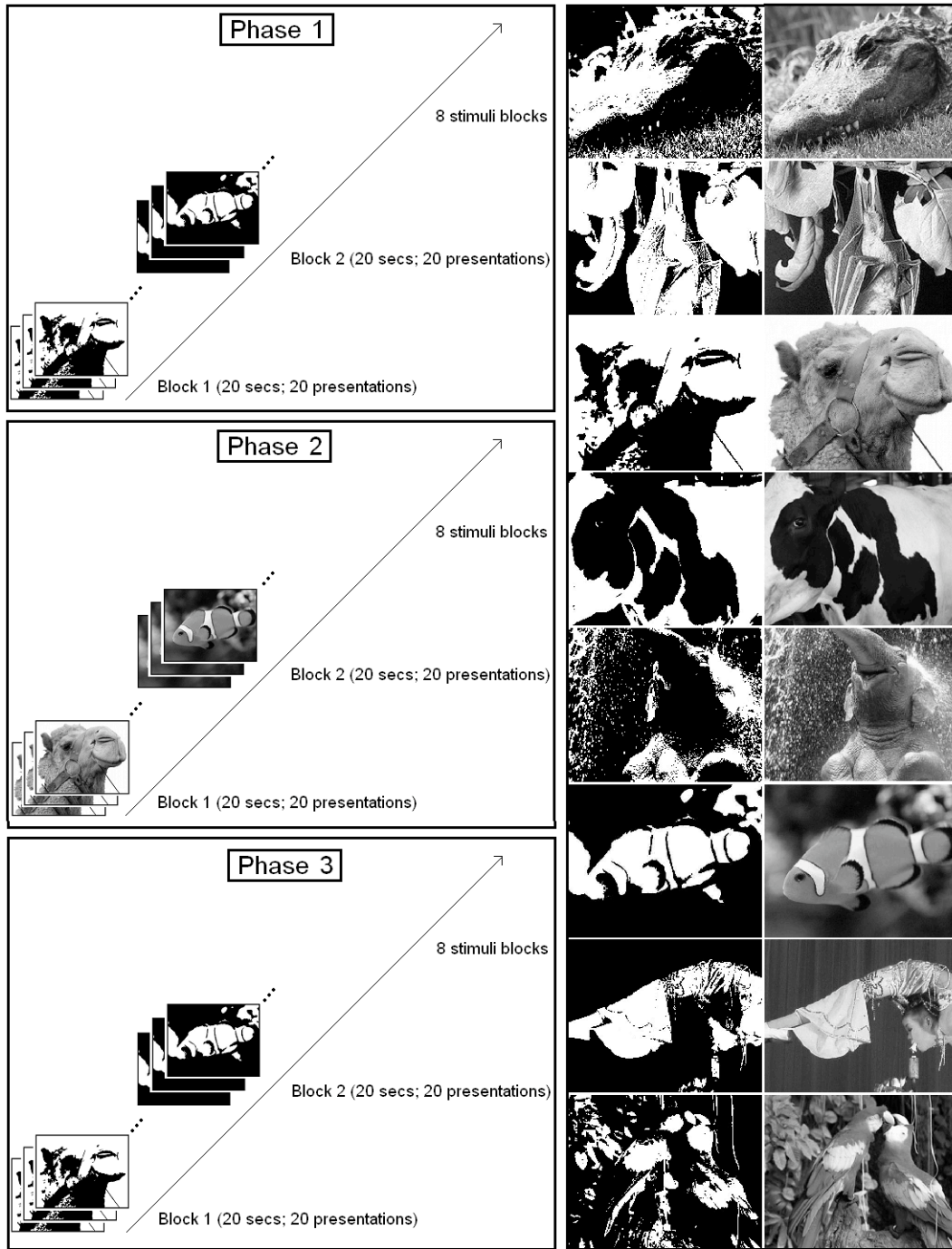
- Bar, M. A cortical mechanism for triggering top-down facilitation in visual object recognition. *J Cogn Neurosci* 15, 600-609 (2003).
- Bar, M., Kassam, K.S., Ghuman, A.S., Boshyan, J., Schmidt, A.M., Dale, A.M., Hamalainen, M.S., Marinkovic, K., Schacter, D.L., Rosen, B.R., & Halgren, E. Top-down facilitation of visual recognition. *Proceedings of the National Academy of Science*, 103, 449-454 (2006).
- Boyaci, H., Fang, F., Murray, S. O., & Kersen, D. Response to lightness variations in early human visual cortex. *Current Biology* 17, 989-993 (2007).
- Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience* 3, 201-215 (2003).
- Datta, R., & DeYoe, E. A. I know where you are secretly attending! The topography of human visual attention revealed with fMRI. *Vision Res.* 49(10):1037-44. (2009).
- Dolan, R. J. et al. How the brain learns to see objects and faces in an impoverished context. *Nature* 389, 596-599 (1997).
- Deshpande, G., Hu, X., Stilla, R., Sathian, K. Effective connectivity during haptic perception: A study using Granger causality analysis of functional magnetic resonance imaging data. *NeuroImage* 40 4, 1807-1814 (2008).
- Dougherty, et al. Visual field representations and locations of visual areas V1/2/3 in human visual cortex, *Journal of Vision* 3, 586–598 (2003).
- Fischer, J. & Whitney, D. Attention narrows position tuning of population responses in V1. *Current Biolgy* 19(16):1356-1361. (2009).
- Fang, F, Kersten, D., Murray, S. O. Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *J Vis.* 8(7):2.1-9. (2008a).
- Fang, F., Boyaci, H., Kersten, D., Murray, S. O. Attention-dependent representation of a size illusion in human V1. *Curr Biol.* 18(21): 1707-1712. (2008b).
- George, N. et al. Contrast polarity and face recognition in the human fusiform gyrus. *Nat Neurosci.* 2, 574-580 (1999).
- Haxby, J. V. et al. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425-2430 (2001).
- Heinen, K., Jolij, J., & Lamme, V. A. (2005). Figure-ground segregation requires two distinct periods of activity in V1: a transcranial magnetic stimulation study. *Neuroreport*, 16(3), 1483-1487.

- Huang, X., & Paradiso, M. A. (2008). V1 Response timing and surface filling-in. *Journal of Neurophysiology*, 100(1), 539-547.
- Hochstein, S. & Ahissar, M. View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron* 36, 791–804 (2002).
- Hupé, J. M. et al. Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394, 784-787 (1998).
- James, T. W., Humphrey G. K., Gati, J. S., Servos, P., Menon, R. S., Goodale, M. A. Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia* 40, 1706-1714 (2002).
- Kamitani, Y. & Tong, F. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8, 679-685 (2005).
- Kersten, D., Mamassian, P. & Yuille, A. Object perception as Bayesian Inference. *Annual Review of Psychology* 55, 271-304 (2004).
- Klein, I., Dubois, J., Mangin, J. F., Kherif, F., Flandin, G., Poline, J. B., Denis, M., Kosslyn, S. M., Le Bihan, D. Retinotopic organization of visual mental images as revealed by functional magnetic resonance imaging. *Brain Res Cogn Brain Res*. 22(1):26-31. (2004).
- Knill, D. C. & Richards, W. editors. *Perception as Bayesian inference*. Cambridge: Cambridge Univ Press. (1996).
- Knill, D. C. & Pouget, A. The Bayesian brain: The role of uncertainty in neural coding and computation for perception and action. *Trends in Neuroscience* 27, 712-719 (2004).
- Kourtzi, Z., & Kanwisher, N. Representation of perceived object shape by the human lateral occipital complex. *Science* 293: 1506–1509. (2001)
- Lee, T. S. A Bayesian framework for understanding texture segmentation in the primary visual cortex. *Vis Res* 5, 2643-2657 (1995).
- Lee, T. S. & Mumford, D. Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A Opt Image Sci Vis*. 20, 1434-1448 (2003)
- Malach, R. et al. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl Acad. Sci. USA* 92, 8135-8139 (1995).
- McKeeff, T. J. & Tong, F. The timing of perceptual decisions for ambiguous face stimuli in the human ventral visual cortex. *Cerebral Cortex* 17, 669-678 (2006).
- Mohr, H. M., Linder, N. S., Linden, D. E., Kaiser, J., Sireteanu, R. Orientation-specific adaptation to mentally generated lines in human visual cortex. *Neuroimage*. 47(1):384-91. Epub (2009).
- Mumford, D. On the computational architecture of the neocortex II. *Biol Cybern* 66, 241-251 (1992).
- Murray, S. O., Boyaci, H. & Kersten, D. The representation of perceived angular size in human primary visual cortex. *Nat Neurosci*. 9, 429-434 (2006).
- Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P. & Woods, D.L. Shape perception reduces activity in human primary visual cortex. *Proc. Natl. Acad. Sci. USA* 99, 15164-15169 (2002).
- Rao, R. P. Bayesian inference and attentional modulation in the visual cortex. *Neuroreport* 16, 1843-1848 (2005).
- Ress, D., Backus, B. T. & Heeger, D. J. Activity in primary visual cortex predicts performance in a visual detection task. *Nat Neurosci*. 3, 940-945 (2000).

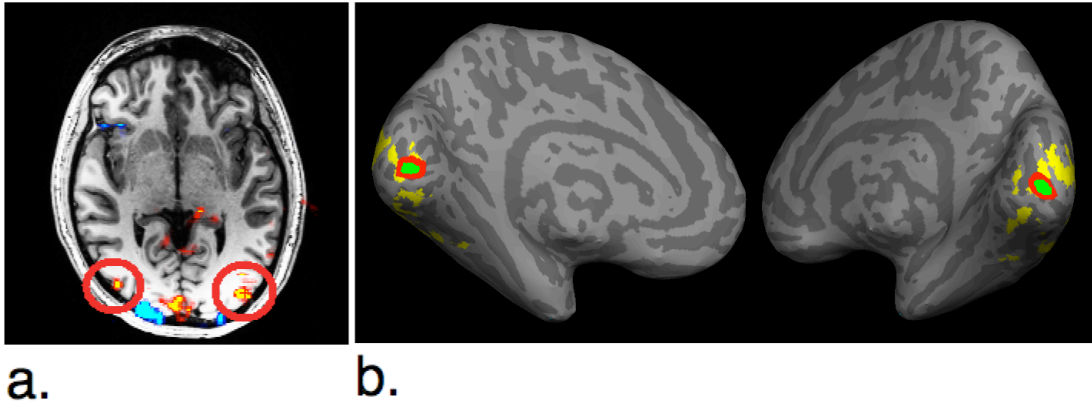


- Roelfsema, P. R., Lamme, V. A. & Spekreijse, H. Object-based attention in the primary visual cortex of the macaque monkey. *Nature* 395, 376-381 (1998).
- Serences, J. T. & Boynton, G. M. Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron* 55, 301-312 (2007).
- Serences, J. T. & Boynton, G. M. The representation of behavioral choice for motion in human visual cortex. *J Neurosci.* 27, 12893-12899 (2007).
- Serences, J. T. Value-based modulations in human visual cortex. *Neuron* 60, 1169-1181 (2008).
- Shuler, M. G. & Bear, M. F. (2006) Reward timing in the primary visual cortex. *Science* 311(5767), 1606-1609.
- Stanley, D. A. & Rubin, N. fMRI activation in response to illusory contours and salient regions in the human lateral occipital complex. *Neuron* 37, 323-331 (2003).
- Ullman, S. Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cereb Cortex* 1, 1-11 (1995).
- Williams, M. A. et al. Feedback of visual object information to foveal retinotopic cortex. *Nat Neurosci.* 11, 1439-1445 (2008).

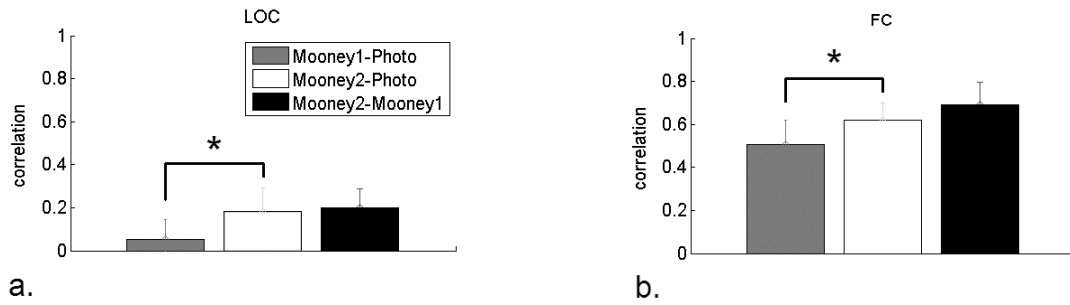
**FIGURE LEGENDS**



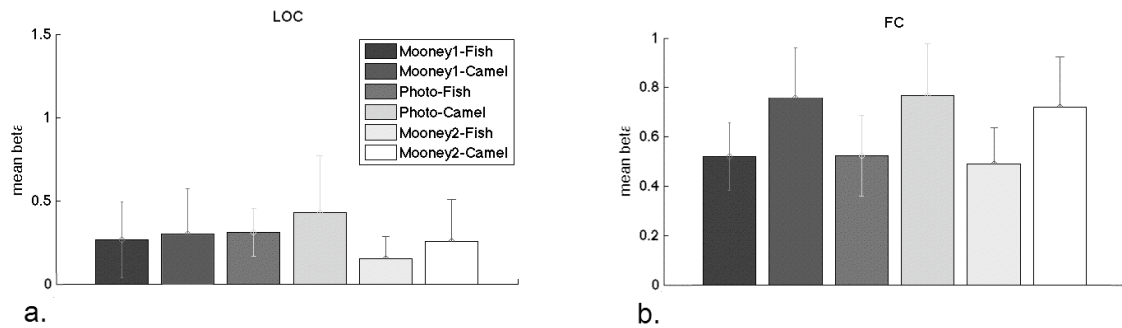
**Figure 1. Experimental procedure (blocked design).** (a) During three phases of the experiment subjects viewed (i) two different two-tone images that they could not identify (“Mooney1”); (ii) the easily identifiable grey-scale versions of the same images (“Photo”); (iii) the original two-tone images which could now be easily recognized because of experience with the corresponding photograph (“Mooney 2”). (b) In Experiment 2, two images were randomly selected for each subject from the set of eight images shown here.



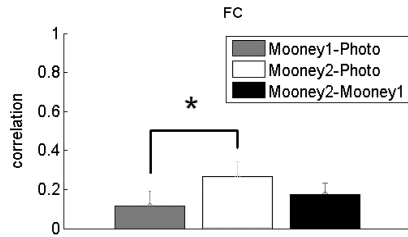
**Figure 2.** Regions of interests. (a) A statistical map for LOC (defined as the region that responded more strongly to images of intact objects than to images of scrambled objects  $p < 10^{-4}$ ) is shown for one representative subject. We selected the two regions at the lateral occipital lobe (circled) as LOC. (b) Inflated cortical surfaces of a brain with dark gray regions representing sulci and light gray regions representing gyri. Foveal confluence (FC) was defined as a small region at the posterior end of the calcarine sulcus (green), functionally constrained to the region (yellow) that responded more strongly to images of intact/scramble objects than to fixation ( $p < 10^{-25}$ ). Another annulus ROI surrounding the FC was also defined (red).



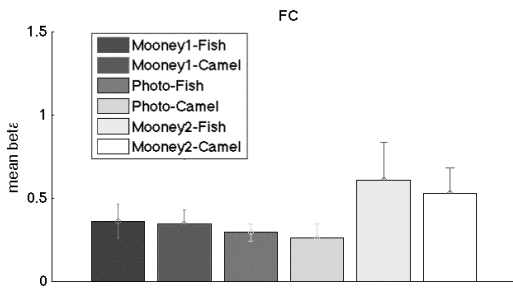
**Figure 3.** Blocked design results. (a) The correlation for the three comparisons (Mooney1-Photo, Mooney2-Photo, and Mooney2-Mooney1) in LOC. The correlation for Mooney2-Photo is significantly greater than for Mooney1-Photo ( $p=0.010$ ). (b) The correlation for the three comparisons in the foveal confluence ROI. The correlation for Mooney2-Photo is significantly greater than for Mooney1-Photo ( $p=0.048$ ). Error bars indicate standard errors.



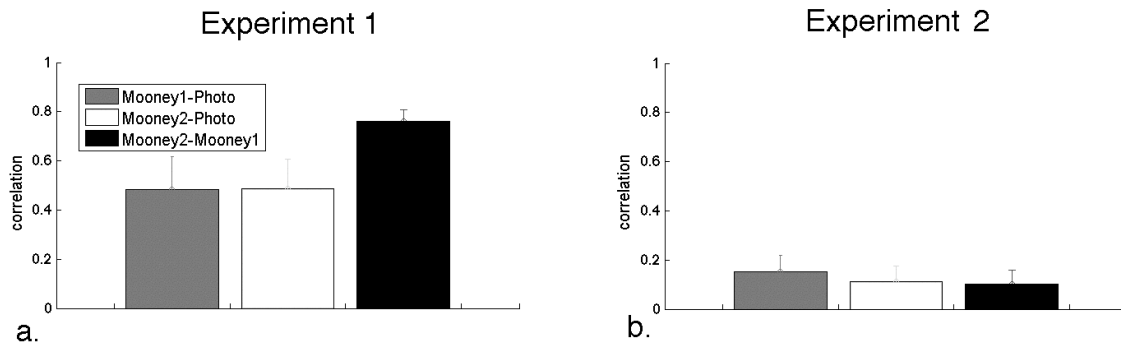
**Figure 4.** The mean beta weights of the 6 conditions for Experiment 1 in (a) LOC and (b) foveal confluence. Error bars indicate standard errors.



**Figure 5.** Event-related results. The correlation for the three comparisons (Mooney1-Photo, Mooney2-Photo, and Mooney2-Mooney1) in the foveal confluence ROI. The correlation for Mooney2-Photo is significantly greater than for Mooney1-Photo ( $p < 0.001$ ). Error bars indicate standard errors.



**Figure 6.** The mean beta weights of the 6 conditions for Experiment 2 in the foveal confluence ROI. Error bars indicate standard errors.



**Figure 7.** Pattern analysis results in the annulus ROI surrounding the foveal confluence. (a) The correlation for the three comparisons (Mooney1-Photo, Mooney2-Photo, and Mooney2-Mooney1) in Experiment 1. The correlation for Mooney2-Photo is not significantly greater than for Mooney1-Photo ( $p = 0.94$ ). (b) The correlation for the three comparisons in Experiment 2. The correlation for Mooney2-Photo is not significantly greater than for Mooney1-Photo ( $p = 0.51$ ). Error bars indicate standard errors.