"Brain-Reading" of Perceived Colors Reveals a Feature Mixing Mechanism Underlying Perceptual Filling-in in Cortical Area V1

Po-Jang Hsieh* **and Peter U. Tse**

Massachusetts Institute of Technology, 46-4141, Cambridge, Massachusetts

◆ ━━━━━━━━━━━━━━━━━━━━━━━━━ ◆

Abstract: Visual filling-in occurs when a retinally stabilized object undergoes perceptual fading. As the term "filling-in" implies, it is commonly believed that information about the apparently vanished object is lost and replaced solely by information arising from the surrounding background. Here we report multivoxel pattern analysis fMRI data that challenge this long-held belief. When subjects view blue disks on a red background while fixating, the stimulus and background appear to turn a uniform purple upon perceptual fading, suggesting that a feature mixing mechanism may underlie color filling-in. We find that ensemble fMRI signals in retinotopic visual areas reliably predict (i) which of three colors a subject reports seeing; (ii) whether a subject is in a perceptually filled-in state or not; and (iii) furthermore, while subjects are in the perceptual state of filling-in, the BOLD signal activation pattern in the sub-areas of V1 corresponding to the location of the blue disks behaves as if subjects are in fact viewing a perceptually mixed color (purple), rather than the color of the disks (blue) or the color of the background (red). These results imply that the mechanism of filling-in in stimuli in which figure and background surfaces are equated is a process of "feature mixing", not "feature replacement". These data indicate that feature mixing may involve cortical areas as early as V1. *Hum Brain Mapp 00:000–000, 2010.*   © 2010 Wiley-Liss, Inc.

Key words: perceptual fading; filling-in; perceptual mixing; feature mixing

◆ ━━━━━━━━━━━━━━━━━━━━━━━━━ ◆

INTRODUCTION

Perceptual fading, also called the "Troxler effect" [Krauskopf, 1963; Troxler, 1804], occurs when an object, though present in the world and continually casting light upon the retina, vanishes from visual consciousness. Perceptual fading occurs most compellingly when the object is located peripherally, has indistinct edges, a low luminance level equal to that of the background, and remains stabilized upon the retina, as happens under conditions of visual fixation [De Weerd, 2006; Livingstone and Hubel, 1987; Komatsu et al., 2000; Spillmann, 2006].

During perceptual "fading," information about the apparently vanished object has traditionally been assumed to be lost and replaced solely by information arising from the surrounding background [Caputo, 1998; De Weerd et al., 1995; Gerrits et al., 1966, 1970; Ramachandran and Gegory, 1991; Spillmann and De Weerd; 2003; Stürzel and Spillmann, 2001; Watanabe and Cavanagh, 1991]. This process of foreground by background feature replacement is commonly known as "filling-in" because the foreground is thought to be filled in by the background feature. In two recent studies [Hsieh and Tse, 2006, 2009], we reported that, when using visual stimuli composed of spatially alternating stripes containing different luminances or motion signals, the filled-in luminance, motion, or color is

approximately the area- and magnitude-weighted average of the background and the foreground luminance, motion, or color, respectively. In light of these data we hypothesized that information within the boundary of the perceptually vanished figure is not lost or replaced by features from outside the boundary during perceptual "filling-in," but instead involves perceptual "feature mixing," whereby information on either side of a perceptually faded boundary merges. In the rest of this article, we use the term "filling-in" to refer to the phenomenology of perceptual changes during Troxler or perceptual fading, but use the terms "feature mixing" and "feature replacement" (or "feature competition") to refer to its possible underling neuronal mechanisms.

Where does perceptual mixing occur in the visual processing pathway? Our goal is to find visual areas where feature information carried by patterns of voxel activation correspond to visual experience. Specifically, we used functional magnetic resonance imaging (fMRI) to carry out a multivoxel pattern analysis (MVPA) of blood oxygenation-level dependent (BOLD) signal activation in early visual cortical areas. It has been previously shown that MVPA can be used to "brain read" cortical selectivity to orientation and motion [Kamitani and Tong, 2005, 2006], orientation and color perception during binocular rivalry [Haynes and Rees, 2005a,b], face and object perception [Haxby et al., 2001], and even intention [Haynes et al., 2007]. In light of this powerful new tool for distinguishing cortical selectivity for different perceptual states and perceived colors, we performed MVPA in early visual cortical areas corresponding to the to-be-filled-in, perceptually vanished figural region, while subjects were in a perceptual state of filling-in. On the basis of our "feature mixing" hypothesis that information within the boundary of the perceptually vanished figure is a mixture of information on either side of a perceptually faded boundary (i.e. a mixture of the foreground and background), we predict that, when subjects are in a perceptual state of filling-in, the BOLD activation pattern in those areas that realize the neural basis of perceptual filling-in will behave as if subjects perceive a mixed color (e.g. purple if the foreground is blue and the background red). Alternatively, according to the "feature replacement" hypothesis, there might be some type of winner take all competition among features of the two surface areas.

METHODS

Participants

All eight (five male, three female; age, 24–46) subjects had normal or corrected-to-normal vision and, prior to participating, gave written, informed consent according to the guidelines of the Department of Psychological and Brain Sciences, and the IRB of Dartmouth College. Subjects received $20 for their participation in each MRI scanning session.

Stimulus Presentation fMRI

Session one

The following psychophysical data were collected in the fMRI scanner. Subjects were presented with full-screen colors (red, blue, and purple; blocked design). Before the fMRI experiment started, the luminance of the red and blue colors were calibrated via psychophysical adjustment (via the technique of Anstis and Cavanagh, 1983; see below for methods) to equal the luminance of the gray background. This was done individually, for each subject. The purple color was then obtained as follows: We presented subjects with three blue disks on a red background, all confined to the upper half of the screen (Fig. 1a). Each blue disk subtended 2° in radius, with a linearly blurred contour extended to 4° in radius. The three disks were centered at 6° above the fixation spot, and 0, +9.5, and −9.5° away from the vertical meridian respectively. Subjects were asked to fixate until perceptual filling-in occurred over all disks, and were then required to adjust the color of the lower half of the screen to match the perceptually filled-in color when they were in the "uniform field" perceptual state. The measured color was then fixed and applied to the purple blocks for the duration of the experiment.

To assure fixation and wakefulness, subjects were required to fixate on the fixation spot and press one of two buttons (with their right-hand index finger switching between two buttons) when the fixation point changed color to red or blue, respectively.

Each subject participated in multiple runs (min = 4, max = 6) in the scanner. Each run lasted 380 s (190 volumes, TR = 2 s), and contained nine 20-second (10 volumes) stimulation blocks. These blocks were interleaved with ten 20-s blank periods during which only the fixation spot was present. Thus each run began and finished with blank periods. In addition, 8 s (4 TRs) of dummy images were collected at the beginning of each run to let the MR signals reach a steady state. In each of the stimulation blocks, one of the three colors was presented (uniform field). In each run, three blocks of each condition were presented. The order in which the stimulus groups were presented was pseudorandomly determined for each run (there were six possibilities: red→blue→purple, red→purple→blue, blue→red→purple, blue→purple→red, purple→red→blue, and purple→blue→red).

Session two

We presented subjects with six blue disks on a red background, which had an identical configuration to the above-described stimuli used for obtaining the perceived purple color but duplicated to cover both the upper and lower halves of the screen (Fig. 1b).

Subjects were required to carry out a dual task: (1) indicating their percepts (filling-in or no-filling-in) by pressing

two buttons with their right-hand index finger, and (2) pressing another button with their left-hand index finger when the fixation point transiently changed color to red. Note that subjects were required to press the "filling-in" and the "no-filling-in" buttons only when the six disks had all disappeared or were all visible. Thus there were durations where subjects pressed neither the "filling-in" nor the "no-filling-in" buttons.

The stimuli were presented for 170 TRs, separated by 10 TRs of blanks in the beginning and end of each run, during which only the fixation task was presented on a gray background. Each subject participated in three runs.

Session three

All stimuli and experimental procedures were identical to session one, except that the luminance of the red and blue colors was subjectively equal to the individually specified purple color experienced during perceptual filling-in (see below for methods details). All stimuli were presented in only the right eye of a binocular goggle system (30 × 20 degree FoV, refresh rate 60 Hz) (Resonance Technology, Northridge, CA) to avoid any possibility of binocular rivalry.

The Determination of Subjective Isoluminance

For session one, the luminance of the red and blue colors was adjusted to be subjectively equal to the gray background for each subject using the minimal flicker technique [Anstis and Cavanagh, 1983]. Before the imaging data were collected, we presented four red flashing (∼30 Hz) disks located at the same location as those disks shown in Figure 1c. Subjects adjusted the red component of the square's color until minimal subjective flicker was reported. The color of the disks was then fixed and applied to the stimuli for the duration of the experiment. This was carried out in the scanner to ensure that isoluminance was optimized for the actual experimental conditions that followed.

For session three, the luminance of the red and blue color was adjusted to be subjectively equal to the purple color (obtained from session one) using the same minimal flicker technique.

Fixation Task

Eye-movements, wakefulness, and attention to the fovea were controlled for by requiring subjects to perform a fixation task in which subjects had to press buttons to a pseudorandomly occurring change in fixation point color at the center of the screen, far from the blue disks. In sessions one and three, the fixation point (0.3 × 0.3 degree) changed color from yellow to red, or yellow to blue, approximately every 2 TRs (the color stayed red or blue for 500ms and changed back to yellow). Subjects were required to press two buttons respectively (with their right-hand index finger

switching between two buttons) when the fixation point changed color to red or blue. In session two, the fixation point changed color only from yellow to red. Subjects were required to press one button (with their left-hand index finger). The color changes occurred an equal number of times during each block, and in a manner uncorrelated with the color of the rest of the stimulus. This task could only be carried out successfully if subjects fixated during both condition and fixation-only blocks, and attended to the fixation point carefully. Runs that had a correct response rate on the fixation task below 70% were excluded from BOLD signal pattern classification analysis.

Data Acquisition

T1-weighted anatomical images were acquired using a high-resolution 3D magnetization-prepared rapid gradient echo sequence (MPRAGE; 160 sagital slices, TE = 4.6 ms, TR = 9.9 ms, flip angle = 8°, 0.938 × 0.938 × 1 mm^3 voxels) as well as a T1-weighted coplanar anatomical image with the same slice orientation as the EPI (echo planar functional image) data, which was used for coregistration. Continuous whole-brain BOLD signal was acquired at the Dartmouth Brain Imaging Center on a Philips 3T scanner using a standard head coil. Standard gradient-echo T2*-weighted SENSE EPIs sensitive to BOLD contrast were collected using 32 slices (3.5 mm thickness and 3 × 3 mm^2 in-plane voxel resolution, interslice distance 0.5 mm, TR = 2000 ms, T2* echo time = 35 ms, flip angle = 90°, FOV (ap, fh, rl) = 240 × 127.5 × 240 mm^3, ascending interleaved slice acquisition, matrix size = 80 × 80) oriented approximately along the anterior- and posteriorcommissure plane. These slices were sufficient to encompass the entire brain of each subject.

fMRI Data Preprocessing

Data were processed using BRAIN VOYAGER (BV) QX 1.8 and MATLAB software developed in house. Effects of small head movements during and between runs were removed using BV's motion correction algorithms. Slice scan time correction was carried out to correct for the fact that slices were not collected at the same time and were collected in interleaved and descending order. Functional data were not smoothed in the space domain, but any low-frequency temporal fluctuations whose wavelength was greater than 30 TRs were removed. This did not introduce correlations between a voxel and its neighbors. For each subject, the functional data were coregistered to the high-resolution anatomical image without spatial normalization. For each voxel, the resulting time course of each run was z-score normalized.

Retinotopic Mapping

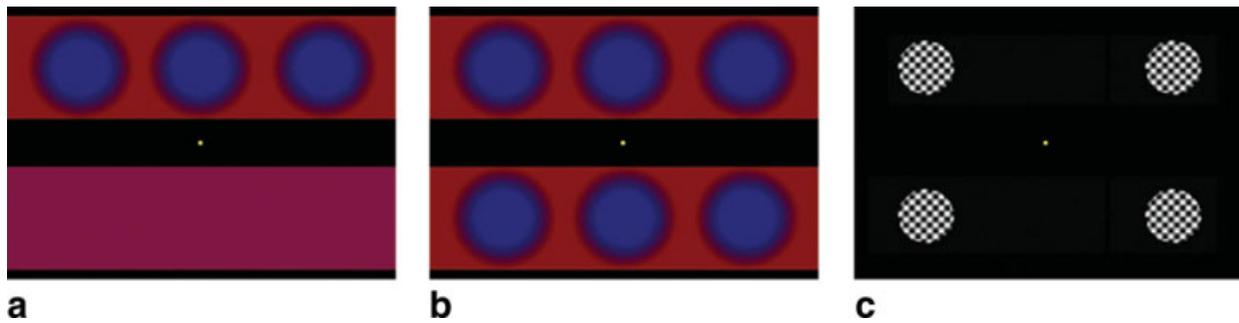Retinotopy (Fig. 2a) was carried out on each subject who participated in the study using standard phase-

Figure 1.

(**a**) The configuration of the stimuli used for finding subjects' perceived uniform color following the onset of perceptual filling-in. We presented subjects with three blue disks on a red background; all were confined to the upper half of the screen. Subjects were asked to fixate until perceptual filling-in occurred, and were then required to adjust the color and luminance of the lower half of the screen to match the perceptually filled-in color and luminance when they were in the "uniform field" perceptual state. (**b**) The configuration of the stimuli used in session two. (**c**) The configuration of the stimuli used for localizing the sub-areas (to-be-filled-in areas) within retinotopic areas that correspond to the four blue disks (in the center of the four quadrants) used in session two. [Color figure can be viewed in the online issue, which is available at www.interscience. wiley.com.]

encoding techniques [Sereno et al., 1995, Slotnick and Yantis, 2003] that we have reported before [Hsieh et al., 2006]. Identified retinotopic areas include V1, V2, V3, and V4v/ VO1 [Brewer et al., 2004].

Constructing Region of Interest Masks

We performed another localization scan to further identify the sub-areas (i.e. the to-be-filled-in areas) within
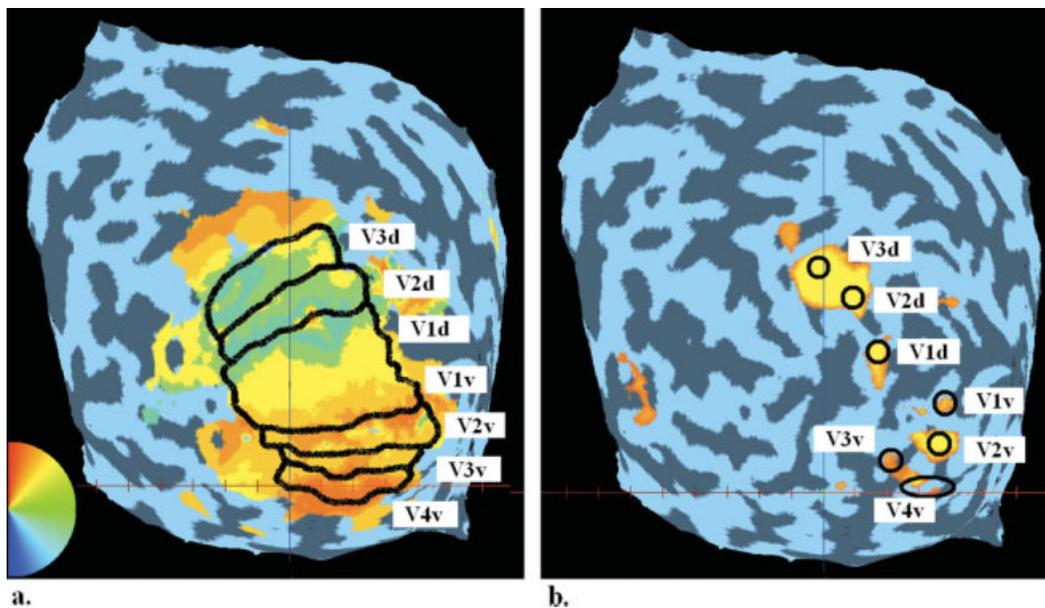

Figure 2.

(**a**) A typical retinotopic map of the flattened left hemisphere occipital pole for one subject is shown with the approximate borders between the retinotopic areas specified in black. Retinotopic area masks were individually specified for each hemisphere of each subject. Blue here represents the lower vertical meridian, cyan/green the horizontal meridian, and red the vertical meridian. (**b**) The sub-areas within the retinotopy that correspond to the to-be-filled-in areas (the center of the four blue disks in the four quadrants). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

retinotopic areas that corresponded to the blue disks used in session two. To avoid any possible unwanted border effect that might occur in the neighborhood of the vertical meridian, we only localized the four disk areas in the center of the four quadrants (Fig. 1c), even though six disks were used in the actual stimulus of blue disks on a red background. Information from the two middle disks was discarded. This localization scan was performed in a block design. Each subject participated in 1-3 runs. Each run lasted 380 s (190 volumes, TR = 2 s), and contained nine 20-s (10 volumes) stimulation blocks. These blocks were interleaved with ten 20-s blank periods during which only the fixation spot was present. Thus each run began and finished with blank periods. In addition, 8 s (4 TRs) of dummy images were collected at the beginning of each run to let the MR signals reach a steady state. In each of the stimulation blocks, four white/black polar checkerboards were presented with a flicker rate of ~2 Hz. Note that centers of these checkerboards were identical to the centers of the four blue disks (in the four quadrants) used in session two.

To restrict the size of these sub-areas to guarantee that they would correspond strictly to the region of the blue disks, with no input from the red background, we only picked the most highly activated inner voxels (~5 voxels) corresponding to each disk (Fig. 2b). As a result, the total number of voxels was ~20 for each region of interest (ROI), leading to 20 dimensions in the BOLD signal pattern classification space. This low dimensionality was conservative not only because we could be certain that the ROI corresponded only to the centers of the blue disks, it was conservative because classification using MVPA classifiers requires greater distance from the classification boundary for a low number of dimensions than a higher number of dimensions.

Nearest Neighbor Classification

Pattern classification was carried out for each subject individually. The data from session one were used as the training data set for each subject. For each voxel within a ROI, the BOLD time course corresponding to each stimulus block was averaged (from the 4th TR to 10th TR, assuming a hemodynamic delay of 2TRs). As a result, each stimulus block generated an "activation pattern" for each ROI. Each activation pattern can be treated as a point in a multidimensional space, with each dimension corresponding to each voxel in the ROI, where magnitude on a dimension was given by the magnitude of BOLD signal activation level within the corresponding voxel. As a result, all BOLD signal activation patterns for a particular kind of stimulus block formed a cluster in this multidimensional space, where each point in a cluster of such points corresponded to one stimulus block.

The data from session two were used as the data set to be classified. For each voxel within a ROI, the BOLD time course corresponding to each perceptually filled-in state was averaged. We only included percepts that lasted longer than 3TRs in order to be sure that a perceptual state was 'pure' and not due to mistaken button presses or subjective uncertainty about perceptual state. When averaging, the first and last TR of each percept were excluded as well, again in order to guarantee a "pure" perceptual state.

We then classified the data of session two based on the three clusters from the training data set. We used the Nearest Neighbor Classification method [Shakhnarovish et al., 2005], which classifies a test data point to one of the three clusters corresponding to red, blue, and purple. Whichever cluster has a shortest Euclidean distance from its center (mean of all data points within that cluster) to a test data point 'won', such that that test data point would be classified as belonging to the same category as that cluster. In other words, a given test pattern x of unknown class was classified to C_K if the Euclidean distance between that test pattern and the center of the class C_K had a minimum value, i.e.:

$$x \in C_K \text{ iff } D_{\text{Euclidean}}(x, C_K) = \min\{D_{\text{Euclidean}}(x, C_T); T = 1, 2, 3\}$$

in which the $D_{\text{Euclidean}}$ represents the Euclidean distance and C_T represents the center of the training cluster T (the mean location of all the data points of that cluster).

Cross-Validation Classification

A similar classification to the one carried out above was performed within the training data set collected from session one. This analysis was carried out to determine whether brain activation patterns were different when viewing different colors. In this analysis, we took each test pattern (i.e. an "activation pattern" for an ROI generated from a stimulus block) and classified it by using the same formula as described above:

$$x \in C_K \text{ iff } D_{\text{Euclidean}}(x, C_K) = \min\{D_{\text{Euclidean}}(x, C_T); T = 1, 2, 3\}$$

This procedure was conducted using a leave-one-out design, such that the center of a training cluster T would be the mean location of all the data points from one particular category (but not including the test data point). In other words, we would have three "centers" from the three categories (corresponding to the red, blue, and purple blocks), and one test data point. The classification process was repeated with iteration through all the training data so that each pattern served as test data only once. If a given pattern was successfully classified as belonging to the category that it should belong to, it would be counted as a correct classification. For each ROI, the percentage of correct classification was calculated within each subject and then averaged across subjects. We would expect the percentage of correct
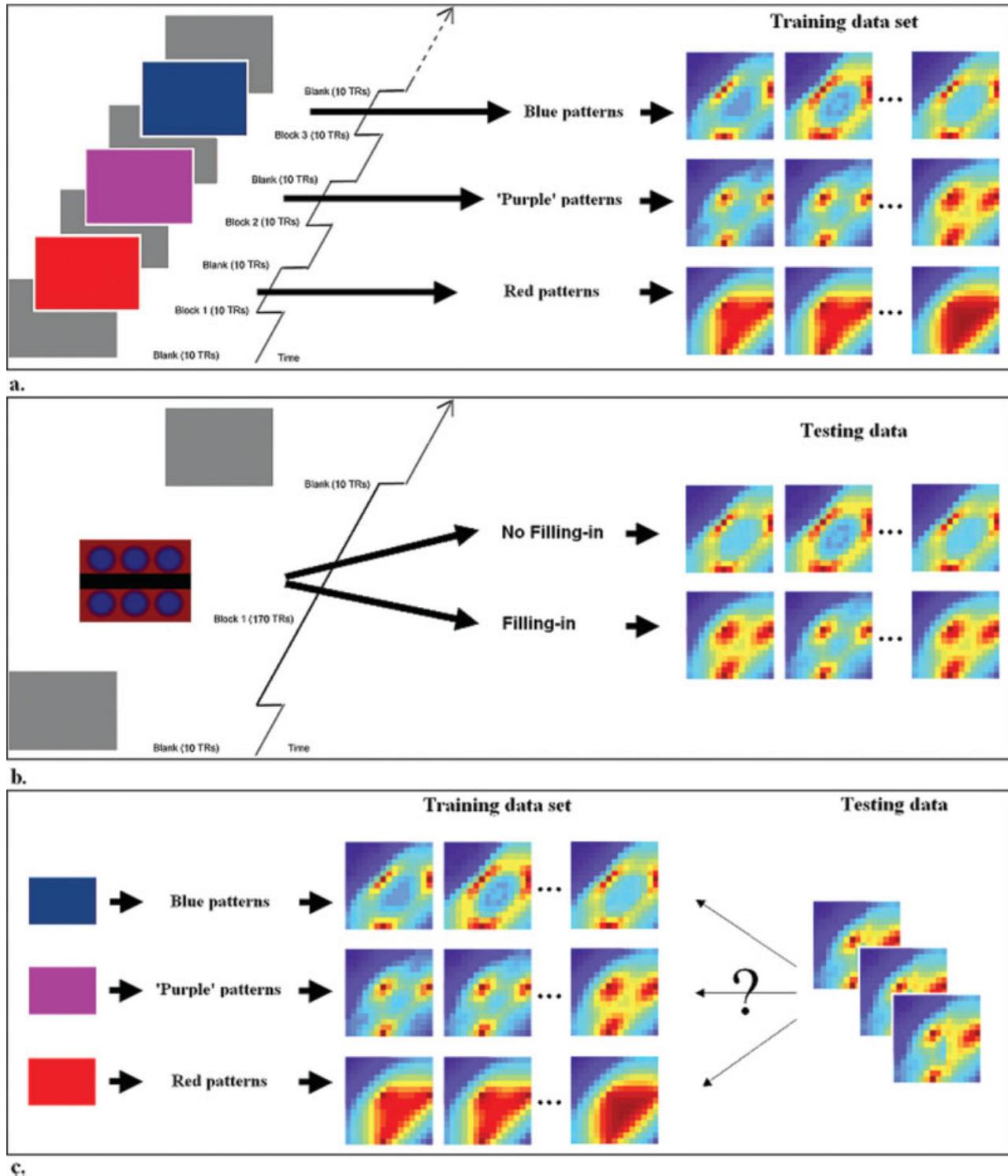
Figure 3.

Stimuli. (**a**) The three sets of BOLD signal activation patterns obtained from session one were used as the training data sets. (**b**) In the second session, we presented six blue disks on a red background, and asked subjects to indicate their dynamic perceptual state (i.e. whether they were in a filling-in state or not at any given time) with a buttonpress. The data were sorted based on subjects' reported precepts. (**c**) We then used multi-voxel pattern analysis to classify the test data from session two as belonging to one of the three possible training data sets. (Note that the patterns of activations in this figure are just symbolic for demonstration purposes.) [Color figure can be viewed in the online issue, which is available at www.interscience. wiley.com.]
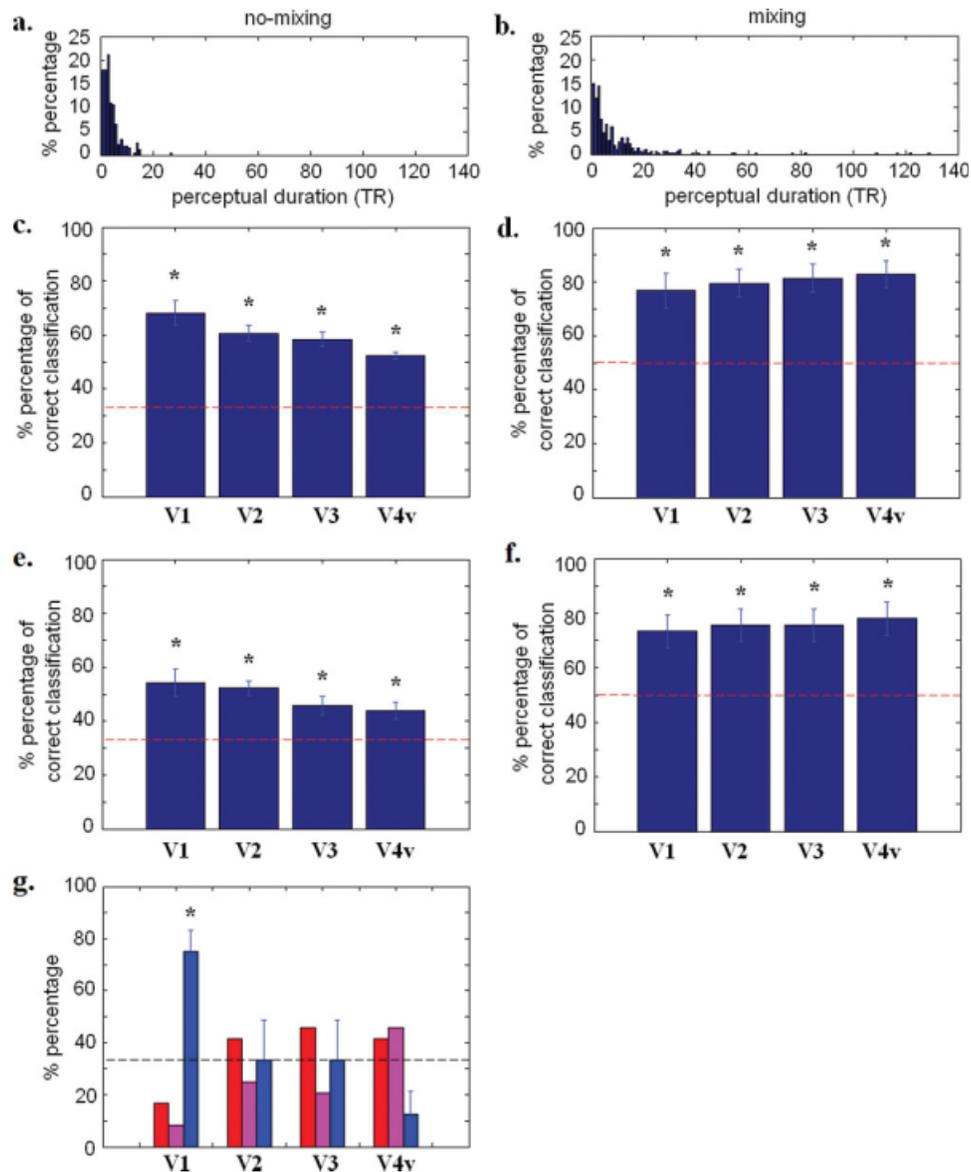
Figure 4.

(a) Histograms of perceptual durations during session two accumulated over eight subjects are shown for the "no-mixing" percept. (b) Histograms of the perceptual durations of the "mixing" percept. (c) When using whole ROIs to classify the training data from session one with cross-validation classification (see Methods), the percentage of correct classification is significantly greater ($P < 0.05$) than the chance rate (33%) in all ROIs (V1, V2, V3, and V4v). (d) When using whole ROIs to classify the data from session two, the brain activation patterns during the filling-in state and no-filling-in state were classified as different clusters. In all ROIs (V1, V2, V3, and V4v), the percentage of correct classification is significantly greater ($P < 0.05$) than the chance rate (50%). (e) When using sub-ROIs to classify the training data from session one, the percentage of correct classification is significantly greater ($P < 0.05$) than the chance rate (33%) in all sub-ROIs. (f) When using sub-ROIs to classify the data from session two, the brain activation patterns during the filling-in state and no-filling-in state were classified as different clusters. In all sub-ROIs, the percentage of correct classification is significantly greater ($P < 0.05$) than the chance rate (50%). Error bars indicate standard error across subjects. Dashed line indicates chance rate. (g) The three color bars in each ROI represent the percentage (averaged across subjects) of the "first-see" states that are classified as belonging to each of the three possible categories (red, blue, or purple). In V1, the percentage of "first-see" states that were classified as belonging to the blue category is higher than the chance rate ($P = 0.0016$), implying that "first-see" states behave more like blue percepts, as expected. Error bars indicate standard error across subjects. The dashed line indicates the chance rate (33.3% because there are three color categories). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

classification to be at the 33% chance rate (chance rate is 33.3% because the test data can be classified to three possible categories) if there is no difference among the three categories.

Similarly, the same cross-validation classification was performed within the data set collected from session two.

Statistics

For each ROI (V1, V2, V3 and V4v/VO1) of each subject, we calculated the percentage of all his/her filling-in percepts that were classified to each of the three possible categories (red, blue, or purple). These numbers were then averaged across subjects for each ROI. A one-sample t-test was performed to test whether they were significantly greater than the 33% chance rate (chance rate is 33.3% because the test data can be classified to three possible categories).

For determining whether two clusters (say clusters A and B) were statistically different, we computed (i) the mean Euclidean distance from each individual data point in cluster A to its center, and (ii) the mean Euclidean distance from each individual data point in cluster B to the center of cluster A. A two-tailed paired t-test was then performed to test whether they were statistically different across subjects.

For the cross-validation classification, we calculated the percentage of correct classification for each ROI (V1, V2, V3, and V4v/VO1) of each subject. A one-sample t-test was performed to test whether they were significantly greater than the chance rate.

Correlation Analysis

To test the possibility that the classification results are driven purely by differences in the mean response but not differences in activation pattern across voxels we repeated the classification process by calculating the correlation coefficient between conditions but not their Euclidean distance. For each subject, correlation coefficients were computed between the activation pattern of each filling-in percept and the mean activation pattern of the red (and the purple) percepts in sub-V1 (using the data from session three). The mean correlation coefficients for these two conditions were than computed within subject and then across subjects. A two-tailed paired t-test was performed to test whether the correlation coefficient was greater between the filling-in percepts and the purple percepts than between the filling-in percepts and the red percepts.

RESULTS

In the first session, we presented subjects with three different uniform colors (red, blue, and purple) spanning the projection screen inside the MRI magnet in a block design.

The three sets of BOLD signal activation patterns obtained were used as the training data to define pattern classifiers (Fig. 3a). In the second session, we presented six blue disks on a red background, and asked subjects to indicate their perceptual state (i.e. whether they were in a filling-in state or not) (Fig. 3b). We then classified these data based on the training data, using multi-voxel pattern analysis (Fig. 3c) (see Methods for details). Note that for each subject, the two colors, red and blue, were individually equated to be subjectively equiluminant, and the uniform purple used was obtained by letting subjects adjust a uniform color field so that it matched the perceived color of the blue disks on a red background after they had shifted into the perceptual state associated with filling-in, namely a uniform field of color upon perceptual fading of the blue disks (see Methods for details). Histograms of perceptual durations during session two are shown in Figure 4a,b, which can be approximated by a gamma distribution[1]. Accuracy of the fixation task averaged across runs was $93.69 \pm 6.24\%$ (runs that had a correct response rate on the fixation task below 70% were excluded).

Color and Perceptual State Decoding

The first analysis was carried out within region-of-interests (ROIs) to classify the training data from session one with cross-validation classification (see Methods). By doing so, we could determine whether brain activation patterns were different when viewing different colors. In all ROIs (V1, V2, V3, and V4v), the percentage of correct classification is significantly greater ($P < 0.0001$) than the chance rate (33%) (Fig. 4c; cumulative decoding plot and control analysis with shuffled labels are shown in Supporting Information Fig. 1). Similarly, when using ROIs to classify the data from session two with cross-validation classification (see Methods), we also found that the brain activation patterns during the filling-in state and no-filling-in state were different. The results show that, in all ROIs (V1, V2, V3, and V4v), the percentage of correct classification is significantly greater ($P < 0.0044$) than the chance rate (50%) (Fig. 4d).

The same pattern classification of BOLD signal was then carried out within sub-regions of V1 ("sub-V1"), V2 ("sub-V2"), V3 ("sub-V3"), and V4v ("sub-V4v") corresponding to the areas occupied solely by the blue disks when they were visible (localized by a localizer scan; see Methods for details). The results show that, in all sub-ROIs, the percentage of correct classification is significantly greater ($P < 0.0119$) than the chance rate (33%) (Fig. 4e). Similarly, when using sub-ROIs to classify the data from session two with cross-validation classification, we also found that the brain activation patterns during the filling-in state and

---

[1]The distribution may look like an exponential distribution. This can happen whenall the short percepts (shorter than 2 secs) are binned into the first data point.

no-filling-in state were different. In all sub-ROIs, the percentage of correct classification is significantly greater ($P < 0.0070$) than the chance rate (50%) (Fig. 4f). Note that when excluding the first perceptual report from the no-filling-in states (these are called the "first-see" states, specified to include the first 7 brain volumes collected after the stimuli first appeared in session 2), the results remain the same (data not shown). This finding suggests that the fMRI responses were not driven by the physical appearance of the blue targets. Together, these results indicate that brain activation patterns were statistically different (i) when subjects viewed different colors, and (ii) when subjects were in different perceptual states (filling-in vs. no-filling-in state).

To confirm that these sub-ROIs were accurately localized, we carried out a control comparison by classifying the "first-see" brain states with the training data. Since subjects always indicated an initial perception upon first appearance of the blue disks on a red background as not one involving perceptual filling-in, the BOLD signal activation patterns in the corresponding sub-areas in V1 should be the same as those induced by the blue uniform color in session 1, since the subjectively visible blue disks were always initially blue. Our results show that, during these "first-see" states, the BOLD signal activation patterns in the sub-regions of V1 corresponding to the locations of the disks did indeed behave as if the subjects perceived the blue color (the percentage of "first-see" states that were classified as belonging to the blue category is higher ($P = 0.0016$) than the chance rate) (Fig. 4g).

To verify that the BOLD signal activation pattern during the "first-see" state is really "the same" as the BOLD signal activation pattern induced by the blue color, and not some pattern that just happened to be closest to it in the metric space used by the classifier algorithm, we tested (paired t-test) whether the two clusters were statistically different (see Methods), and the results show that they are not ($P = 0.1466$), suggesting that the two kinds of BOLD signal activation patterns are indeed the same. This result suggests that (i) the BOLD signal activation patterns behaved normally before entering the filling-in state, and (ii) our algorithm has the ability to correctly classify a BOLD signal activation pattern in V1 as belonging to the category that it should belong to.

Since only sub-V1, and not sub-V2, sub-V3, or sub-V4v, can distinguish blue in the "first-see" state, we decided to only look in sub-V1 to see whether the filling-in state is more like a "red-state," "blue-state," or a "purple-state". That is, since BOLD signal in the subregions of V2, V3, and V4v corresponding to the location of the blue disks cannot be classified as blue by our algorithm when blue is in fact present, there is no point in examining whether BOLD signal there during the perceptual filling-in state resembles the state when the stimulus is in fact a uniform field of blue, red, or purple. However, since sub-V1 could be classified reliably as blue when blue was present and perceived at the location of the blue disks,

sub-V1 can be used to classify the filling-in state as red, blue, or purple.

Colors Perceived During Filling-In States Were Classified as a Mixed Color

After prolonged viewing of blue disks on a red background while maintaining fixation, subjects switched in and out of the perception of a uniformly colored purple field. BOLD signal activation patterns in sub-V1 could mimic those when subjects viewed either (i) a uniform field of the foreground color (i.e. blue) which would be predicted if V1 follows the stimulus but not the filling-in percept, since the bottom-up input to sub-V1 comes solely from the blue disks, or (ii) a uniform field of the perceptually mixed color (i.e. purple), as would be predicted by a "feature mixing" view of color filling-in, since this is what is perceived during the filling-in state, or (iii) a uniform field of the background color (i.e. red), if red replaces (or wins over) blue as would be predicted by the traditional "feature replacement" or "feature competition" view of color filling-in, or finally, the pattern could be distinct from any of the three just mentioned possibilities.[2]

Results (Fig. 5a) show that measured BOLD signal patterns of activation corresponding to filling-in percepts were classified as belonging to the purple category at a rate higher ($P = 0.0019$) than the chance rate (33.3%). We also did the same Euclidean distance comparison for the "filling-in cluster" and the "purple cluster," which revealed that the two clusters are not statistically different ($P = 0.3068$). In other words, BOLD signal patterns of activation during filling-in behave more like those when the subjects were actually viewing a uniform field of the purple (the perceptually mixed color) that they reported seeing when in the filling-in state, implying that the mechanism of "filling-in" is more consistent with a process of featural mixing than one of featural replacement or featural competition, at least in cortical area V1.

One might argue that this classification result might not be based on the color per se, but might be biased by a difference of luminance between the color purple and the colors red and blue, because the perceived luminance was always lower for the individually specified perceptually mixed color (purple) than for the red and blue comprising the stimulus. To control for this possibility, we ran a third experiment by presenting subjects with three uniform color fields (red, blue, and purple) in a block design again, but this time the luminance of the colors red and blue were adjusted in the scanner by each individual subject before the experiment so that the red and blue used were

---

[2]Note that the third possibility is unlikely because it is not consistent with phenomenology, which involves seeing a uniform purple field during the filling-in state. However, it is still possible that a feature replacement (or feature competition) mechanism might co-exist with a feature mixing mechanism, each realized in different areas of the brain. We therefore still list it here as a possible outcome.
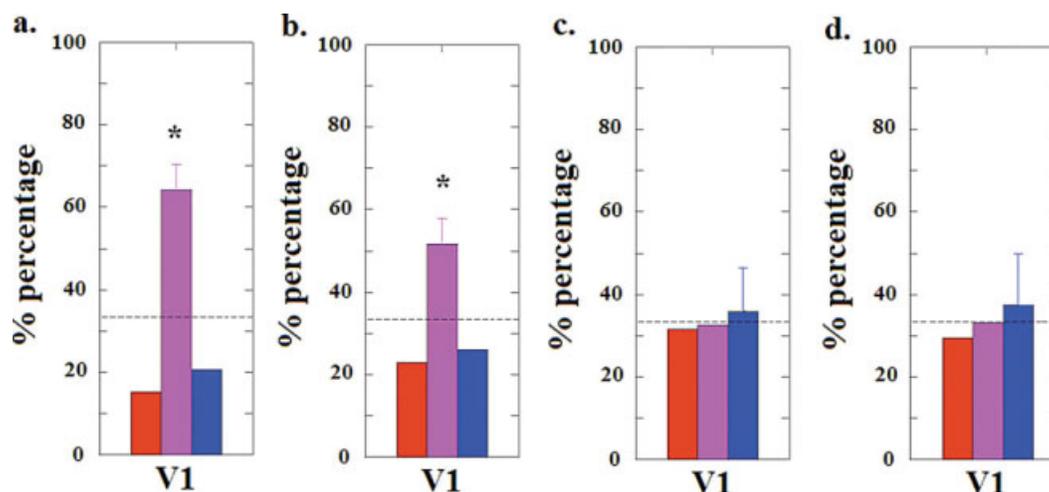
Figure 5.

(**a**) The three color bars represent the percentage (averaged across subjects) of filling-in percepts that were classified as belonging to each of the three respective possible categories (red, blue, or purple) from session 1. The results show that, in sub-V1 (voxels in V1 that correspond strictly to the region of the blue disks), the percentage of filling-in percepts that were classified as belonging to the purple category is higher ($P =$ 0.0019) than the chance rate (33.3%), implying that filling-in percepts behave more like percepts of the color purple than either the color red (the background color) or the color blue (the foreground color). The dashed line indicates the chance rate (33.3%). (**b**) The three color bars represent the percentage (averaged across subjects) of filling-in percepts that are classified as belonging to each of the three possible categories (isoluminant red, blue, or purple) from session 3. The results show that filling-in percepts behave more like purple percepts ($P =$ 0.0130) in sub-V1. (**c**) The three color bars represent the percentage (averaged across subjects) of no-filling-in percepts that were classified as belonging to each of the three respective possible categories (red, blue, or purple) from session 1. The results show that, in sub-V1, the percentage of no-filling-in percepts that were classified as belonging to the blue category is higher than the other two categories though not significantly ($P >$ 0.05). (**d**) The three color bars represent the percentage (averaged across subjects) of no-filling-in percepts that are classified as belonging to each of the three possible categories (equiluminant red, blue, or purple) from session 3. The results show that, in sub-V1, the percentage of no-filling-in percepts that were classified as belonging to the blue category is higher than the other two categories though not significantly ($P >$ 0.05). Error bars indicate standard error across subjects. The dashed line indicates the chance rate (33.3%). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

equiluminant to the color purple specified by subjects in experiment 2. When using the three sets of BOLD signal activation patterns obtained from the third session as the training data sets, our classification results continue to show that the BOLD signal activation pattern in the sub-V1 (corresponding to the locations of the blue disks) behaves as if subjects were perceiving a perceptually mixed (purple) color rather than the background replacement (red) color (the percentage of filling-in percepts that were classified as belonging to the purple category is higher ($P =$ 0.0130) than the chance rate (Fig. 5b). Similarly, the results of a Euclidean distance comparison show that the two clusters are not statistically different ($P =$ 0.4450). This effectively rules out the potential luminance confound.

We also tested whether the no-filling-in percepts from both sessions 1 and 3 were more likely to be classified as belonging to the blue category. The results in Figure 5c,d show that in V1 the percentage of no-filling-in percepts that were classified as belonging to the blue category is only slightly higher than the other two categories (not sig-

nificantly; $P >$ 0.05). We suspect that the no-filling-in percepts might reflect a perceptually faded blue state and therefore cannot be perfectly classified as blue.

Finally, to test the possibility that the classification results are purely driven by differences in the mean response, we repeat the classification process by calculating the correlation coefficient between conditions rather than by their Euclidean distances (see Methods). For each subject, correlation coefficients were computed between the activation pattern of each filling-in percept and the mean activation pattern of the red (and the purple) percepts in sub-V1 (using the data from session three). Results show that (Supporting Information Fig. 3), with a smaller sub-V1 (~20 voxels), the correlation coefficient between the filling-in percepts and the purple percepts is slightly, but not significantly, greater than that between the filling-in percepts and the red percepts ($P =$ 0.1308). However, with a bigger sub-V1 (~40 voxels), the correlation coefficient between the filling-in percepts and the purple percepts is significantly greater than that between the filling-

in percepts and the red percepts ($P = 0.0021$). We suspect that this is because a larger number of voxels are required for an effective correlation analysis. These results suggest that the classification results are not purely driven by differences in the mean response.

DISCUSSION

Several recent neuroimaging studies have shown effects of filling-in in V1, including neon-color spreading (a visual illusion that occurs when the color or lightness of certain physical inducers spreads over an illusory shape; Sasaki and Watanabe, 2004], phantom illusion (an illusory filled-in percept observed between the gap region between two moving gratings; Meng et al., 2005], blind spot filling-in [Tong and Engel, 2001], texture filling-in [Weil et al., 2007, 2008], and Troxler fading [Mendola et al., 2006]. However, when using the Craik-O'Brien-Cornsweet illusion[3] [Cornsweet, 1970; Grossberg and Todorović, 1988], no correlated response was observed in early visual areas [Perna et al., 2005]. Similarly, in a study using brightness/color induction (the modulation of the perceived intensity of a region by the luminance/color of surrounding regions), observed fMRI activity was considered to be unrelated to surface filling-in, but rather was attributed to a linear combination of short-range and long-range responses elicited by luminance/color edges [Cornelissen et al., 2006].

Similar contradictory results also exist in neurophysiological data. For example, several studies have reported effects of blind spot filling-in in V1 [Fiorani et al., 1992; Komatsu, 2006; Matsumoto and Komatsu, 2005]. However, when using the Craik-O'Brien-Cornsweet illusion, neuronal activity that modulated with illusory brightness was only discovered in thin stripes of V2. Similarly, neuronal activity correlated with texture filling-in was found in V2 and V3, but not in V1 [De Weerd et al., 1995], and illusory contours evoke responses mostly in V2 cells [Peterhans and von der Heydt, 1989; von der Heydt et al., 1984 von der Heydt and Peterhans, 1989]. When examining neuronal activity associated with color filling-in in the Troxler effect, von der Heydt et al. [2003] did not observe clear modulation of neuronal activity in V1 and V2 surface cells (neurons in V1 and V2 that responded to uniform surfaces and had receptive fields inside the Troxler target disc), suggesting that surface cells respond to the retinal disc color regardless of whether filling-in occurs or not.

The discrepancy between these different studies using different stimuli and methodologies needs to be resolved. Komatsu has pointed out in a review [Komatsu, 2006] that these differences might reflect (i) differences in the underly-

ing neural mechanisms, (ii) differences in the stimuli used, or (iii) differences in the recording layer (neuronal activities related to blind spot filling-in were recorded in deep layers of V1, but those related to the Craik-O'Brien-Cornsweet illusion and Troxler fading were mostly recorded in the superficial layers of V1). Further studies are therefore required to resolve these contradictory results.

Our finding is consistent with a previous fMRI study on Troxler fading [Mendola et al., 2006; Weil et al., 2008], that found that filling-in may occur as early as in V1. Importantly, our result goes beyond this and suggests that this "filling-in" mechanism may involve feature mixing but not feature replacement or feature competition, where one color wins out over another, and perceptually replaces it. This is also consistent with our recent psychophysical data supporting feature mixing over feature replacement or feature competition as the mechanism underlying altered perception during perceptual fading [Hsieh and Tse, 2006, 2009].

It is worth noting that our current design is limited by the colors we used (blue foreground and red background) and its specific configuration (figure and ground surface areas are about equal). Although we have previously shown with psychophysical methods that feature mixing also occurs over different feature domains such as luminance and motion signal, and different paradigms such as neon-color-spreading [Hsieh and Tse, 2009], future experiments are still required to test whether the neuronal effect we observed here is generalizable over different combinations of colors and figure configurations.

Lastly, it is worth noting that the inability to "read" perceptual states in subareas beyond V1 does not imply that these areas are not involved in perceptual filling-in. It is possible that the small size of the target region leads to a less accurate specification of ROIs in higher areas because of increasing receptive field size. It is also possible that neuronal structure in higher visual areas is less "heterogeneous" than early visual areas (i.e. there is no reported orientation column structure or segregated color blob structure in higher visual areas), making their BOLD activation patterns less biased by random within-voxel differences. This might explain the results in Figure 4c that show a decrease of classification performance for higher visual areas.

To conclude, after prolonged viewing of blue disks on a red background, subjects alternate between a state where they veridically perceive the stimulus and a non-veridical state of perceptual filling-in, where the stimulus appears to be a uniform field of purple. While subjects are in the perceptual filled-in state, the BOLD signal activation pattern in the sub-areas of V1 corresponding to the location of the blue disks behaves as if subjects were in fact viewing a perceptually mixed color (purple), rather than the color of the disks (blue) or the color of the background (red). Thus, the pattern of BOLD signal activity in V1 supports the feature mixing theory over the feature replacement theory or feature competition theory of perceptual filling-in, at least in V1. Based on the data in this study, the possibility can be raised that also in displays in which the figure is much smaller than

---

[3]An illusion where the region adjacent to the lighter side of an edge looks slightly lighter than the region adjacent to the darker side of that edge, but where, in fact, the brightness of both regions is exactly the same.

the background, the underlying process leading to perceptual filling-in reflects a mixing of surface features related to the background, and features related to the figure. According to this idea, observers would report filling-in because the effect of the spreading of the background outweighs the effect of the spreading of the figure, though that spreading may indeed take place. Alternatively, the imbalance between the neural populations encoding the figure and the background may lead to a more competitive process in which the background may win. The present data cannot settle this issue. Future investigations are required to determine whether feature mixing also occurs in more traditional filling-in studies, where the filled-in region is small compared to the extent of the background.

REFERENCES

Anstis SM, Cavanagh P (1983): A minimum motion technique for judging equiluminance. In: Mollon JD, Sharpe LT, editors. Color Vision: Psychophysics and Physiology. London: Academic Press. pp 66–77.

Brewer AA, Liu J, Wade AR, Wandell BA (2004): Human ventral occipitotemporal cortex contains several visual field maps with differential stimulus selectivity. Soc Neurosci Meeting. Abstract 300.23.

Caputo G (1998): Texture brightness filling-in. Vis Res 38:841–851.

Cornelissen FW, Wade AR, Vladusich T, Dougherty RF, Wandell B (2006): No functional magnetic resonance imaging evidence for brightness and color filling-in in early human visual cortex. J Neurosci 26:3634–3641.

Cornsweet T (1970): Visual Perception. New York: Academic.

De Weerd P, Gattass R, Desimone R, Ungerleider LG(1995): Responses of cells in monkey visual cortex during perceptual filling-in of an artifical scotomas. Nature 377:731–734.

De Weerd P (2006): Perceptual filling-in: More than the eye can see. Prog Brain Res 154:227–245.

Fiorani M, Rosa MGP, Gattas R, Rocha-Miranda CE (1992): Dynamic surrounds of receptive fields in primate striate cortex: A physiological basis for perceptual completion? Proc Natl Acad Sci USA 89:8547–8551.

Gerrits HJM, deHaan B, Vendrik AJH(1966): Experiments with retinal stabilized images. Relations between the observations and neural data. Vis Res 6:427–440.

Gerrits HJM, Vendrik AJH (1970): Simultaneous contrast, filling-in process and formation processing in man's visual system. Exp Brain Res 72:279–286.

Grossberg S, Todorović D (1988): Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. Perception Psychophys 43:241–277.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P(2001): Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430.

Haynes JD, Rees G (2005a): Predicting the stream of consciousness from activity in human visual cortex. Curr Biol 15:1301–1307.

Haynes JD, Rees G (2005b): Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci 8:686–691.

Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE. (2007): Reading hidden intentions in the human brain. Curr Biol 17:323–328.

Hsieh P-J, Caplovitz G, Tse PU (2006): Bistable illusory rebound motion: Event-related functional magnetic resonance imaging of perceptual states and switches. Neuroimage 32:728–739.

Hsieh P-J, Tse PU (2006): Illusory color mixing upon perceptual fading and filling-in does not result in 'forbidden colors'. Vis Res 46:2251–2258.

Hsieh P-J, Tse PU (2009): Feature mixing rather than feature replacement during perceptual filling-in. Vis Res 49:439–450.

Kamitani Y, Tong F (2005): Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679–685.

Kamitani Y, Tong F (2006): Decoding seen and attended motion directions from activity in the human visual cortex. Curr Biol 16:1096–1102.

Kanai R, Kamitani Y (2003): Time-locked perceptual fading induced by visual transients. J Cogn Neurosci 1;15:664–672.

Krauskopf J (1963): Effect of retinal image stabilization on the appearance of heterochromatic targets. J Opt Soc Am 53:741–744.

Komatsu H. (2006). The neural mechanisms of perceptual filling-in. Nat Rev Neurosci 7:220–231.

Komatsu H, Kinoshita M, Murakami I (2000): Neural responses in the retinotopic representation of the blind spot in the macaque V1 to stimuli for perceptual filling-in. J Neurosci 20:9310–9319.

Livingstone MS, Hubel DH (1987): Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. J Neurosci 7:3416–3418.

Matsumoto M, Komatsu H (2005): Neural responses in the macaque V1 to bar stimuli with various lengths presented on the blind spot. J Neurophysiol 93:2374–2387.

Mendola JD, Conner IP, Sharma S, Bahekar A, Lemieux S(2006): fMRI Measures of perceptual filling-in in the human visual cortex. J Cogn Neurosci 18:363–375.

Meng M, Remus DA, Tong F (2005): Filling-in of visual phantoms in the human brain. Nat Neurosci 8:1248–1254.

Perna A, Tosetti M, Montanaro D, Morrone MC(2005): Neuronal mechanisms for illusory brightness perception in humans. Neuron 47:645–651.

Peterhans E, von der Heydt R (1989): Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. J Neurosci 9:1749–1763.

Ramachandran VS, Gregory RL (1991): Perceptual filling in of artificially induced scotomas in human vision. Nature 350:699–702.

Sasaki Y, Watanabe T (2004): The primary visual cortex fills in color. Proc Natl Acad Sci USA 101:18251–18256.

Sereno MI, Dale A, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995): Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889–893.

Shakhnarovish G, Darrell T, Indyk P (2005): Nearest-Neighbour Methods in Learning and Vision. Cambridge Mass: The MIT Press.

Slotnick SD, Yantis S (2003): Efficient acquisition of human retinotopic maps. Hum Brain Mapp 18:22–29.

Spillmann L. (2006). From perceptive fields to Gestalt. Prog Brain Res 155:67–92.

Spillmann L, DeWeerd P (2003): Mechanisms of surface completion: Perceptual filling-in of texture. In: Pessoa L, DeWeerd P, editors. Filling-In: From Perceptual Completion to Cortical Reorganization. London: Oxford University Press. pp 81–105.

Stürzel F, Spillmann L (2001): Texture fading correlates with stimulus salience. Vis Res 41:2969–2977.

Tong F, Engel SA (2001): Interocular rivalry revealed in the human cortical blind-spot representation. Nature 411:195–199.

Troxler D (1804): Über das Verschwinden gegebener Gegenstände innerhalb unsers Gesichtskreises. In: Himly K, Schmidt JA, editors. Ophthalmologische Bibliothek II. Jena: Fromman. pp 51–53.

von der Heydt R, Peterhans E, Baumgartner G (1984): Illusory contours and cortical neuron responses. Science 224:260–262.

von der Heydt R, Peterhans E (1989): Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. J Neurosci 9:1731–1748.

von der Heydt R, Friedman H, Zhou H (2003): In: Pessoa L, De Weerd P, editors. Filling-in. New York: Oxford University Press. pp 106–127.

Watanabe T, Cavanagh P (1991): Texture and motion spreading, the aperture problem, and transparency. Percept Psychophys 50:459–464.

Weil RS, Kilner J, Haynes JDH, Rees G (2007): Neural correlates of perceptual filling-in of an artificial scotoma in humans. Proc Natl Acad Sci USA 104:5211–5216. ISSN: 0027–8424.

Weil RS, Watkins S, Rees G (2008): Neural correlates of perceptual completion of an artificial scotoma in human visual cortex measured using functional MRI. NeuroImage 42:1519–1528.