Neural circuits for long-range color filling-in

Peggy Gerardin[1], Clément Abbatecola[1], Frédéric Devinck[2], Henry Kennedy[1], Michel Dojat[3], Kenneth Knoblauch[1]

[1] Univ Lyon, Université Claude Bernard Lyon 1, Inserm, Stem Cell and Brain Research Institute U1208, 69500 Bron, France.

[2] Université Rennes 2, CRPCC, EA 1285, 35000 Rennes, France

[3] Univ. Grenoble Alpes, Inserm, CHU Grenoble Alpes, GIN, 38000 Grenoble, France

Running Title: Neural correlates of filling-in.

Correspondence should be addressed to: Peggy Gerardin & Kenneth Knoblauch, INSERM U1208 Stem-Cell and Brain Research Institute, 69500 Bron, France.

Email: peggy.gerardin@inserm.fr, ken.knoblauch@inserm.fr

Abstract

Surface color appearance depends on both local surface chromaticity and global context. How are these inter-dependencies supported by cortical networks? Combining functional imaging and psychophysics, we examined if color from long-range filling-in engages distinct pathways from responses caused by a field of uniform chromaticity. We find that color from filling-in is best classified and best correlated with appearance by two dorsal areas, V3A and V3B/KO. In contrast, a field of uniform chromaticity is best classified by ventral areas hV4 and LO. Dynamic causal modeling revealed feedback modulation from area V3A to areas V1 and LO for filling-in, contrasting with feedback from LO modulating areas V1 and V3A for a matched uniform chromaticity. These results indicate a dorsal stream role in color filling-in via feedback modulation of area V1 coupled with a cross-stream modulation of ventral areas suggesting that local and contextual influences on color appearance engage distinct neural networks.

Keywords: Functional MRI, Surface Color Appearance, Human Vision, Functional Streams, Cortical Hierarchy

1. Introduction

Surface color appearance depends on multiple factors. In the absence of contextual information, color appearance largely depends on surface spectral information and will be referred to as *surface-dependent color*. Nevertheless, contextual features influence and modulate color appearance. Surface color perception that is driven by remote contours is referred to as filling-in. Long-range filling-in phenomena suggest that the color appearance of surfaces depends on combining local surface information with contextual information, such as from distant edges. We will refer to the latter as *edge-dependent or induced* color. The contributions of contextual influences to surface color appearance have interested scientists for many years and have been successfully studied via a number of visual phenomena, including simultaneous contrast (Chevreul 1839), neon-color spreading (Varin 1971), and, more recently by the Watercolor Effect (Pinna, Brelstaff, and Spillmann 2001). In contrast to the progress made on the psychophysics of these phenomena, little is known about the distributed neural representation of edge-induced filling-in percepts and to what extent identical percepts induced by edge and surface information depend on common or distinct neural networks.

A common hypothesis is that surface- and edge-dependent percepts generate equivalent neural activity at early visual stages (Komatsu 2006). Ratliff and Sirovich (1978) noted that convolution with center-surround like filters both of step-edges and of the spatial transients that induce filling-in (e.g., Craik-O'Brien-Cornsweet effects) results in nearly identical images (Figure 1). One hypothesis is that the center-surround receptive field organization in early vision could lead to edge transients and surfaces generating equivalent neural response profiles that the observer would perceive similarly.

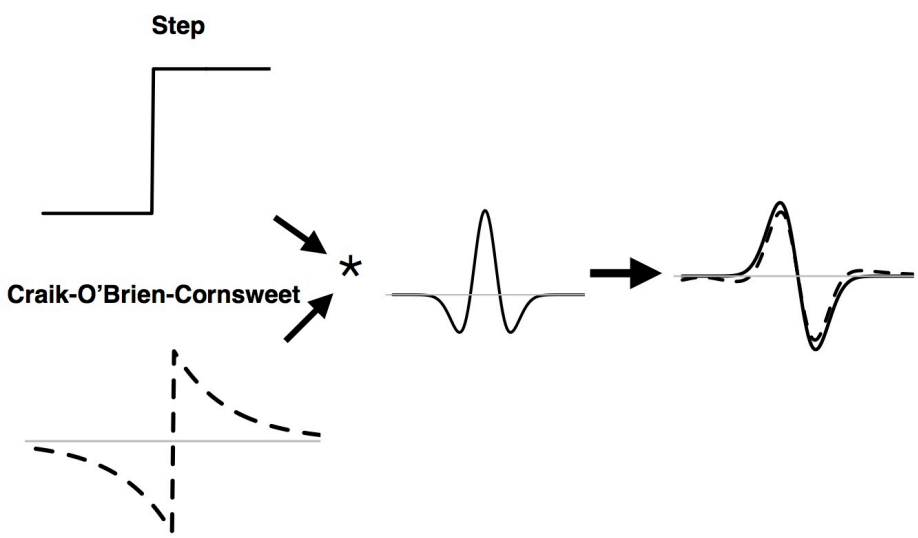

Figure 1: Equivalence of responses for step and gradient intensity profiles under convolution with a center-surround weighting function.

Both a step change in intensity (solid curve, upper left) and the exponential ramp transient for a Craik- O'Brien-Cornsweet effect (dashed curve, lower left) when convolved (indicated by "*") with a center-surround weighting function (center, the negative second derivative of a Gaussian) yield nearly identical response profiles (right, solid and dashed curves, respectively).

Bayesian models of perception predict that both such stimuli would be perceived as uniform fields because they would correspond to the most likely cause of the neural activation (Brown and Friston 2012). This proposal is agnostic, however, with respect to whether the underlying neural representations generate retinotopically-distributed activity isomorphic with the fill-in percept at a subsequent stage (Anstis 2010). Such a proposal is compatible with reports that in cortical area V1 the most frequently encountered color sensitive cells are double-opponent that respond best to isolated chromatic contours but also to the edges of uniform color and luminance surfaces (Friedman, Zhou, and von der Heydt 2003, Johnson, Hawken, and Shapley 2001). These cells show response profiles that are stronger for stimulus edges than uniform surfaces (Friedman, Zhou, and von der Heydt 2003), and, in fact, optical imaging experiments using voltage sensitive dyes in macaque (Zweig et al. 2015) and functional imaging in humans (Cornelissen et al. 2006) demonstrate that the response profiles in area V1 for uniform luminance or chromatic surfaces are dominated by edge responses.

In area V1, it is conceivable that surface- and edge-dependent percepts could be processed through distinct neural channels. Double-opponent cells in area V1 could carry information about both isolated chromatic and luminance edge-transients, such as those that generate filling-in, as well as the edges of uniform surfaces while single-opponent cells would respond optimally to the interior region of uniform fields. Single-opponent receptive fields display different spectral sensitivities in spatially distinct excitatory and inhibitory regions, thereby giving rise to spectrally dependent responses to uniform fields. Such an organization would be consistent with theories of visual processing that propose independent surface and edge processing in form perception (Pinna and Grossberg 2005). While the population of single opponent cells in V1 is estimated to be less than half that of the double-opponent cells, this may simply reflect differences in the sampling requirements for detecting uniform surfaces and edges (Schluppeck and Engel 2002). In higher visual areas little is known about the relative contribution of edge-dependent and local surface processing mechanisms. Here, we ask whether or not the neural activity resulting from edge-induced color filling-in of a surface is channeled through the same cortical inter-areal networks as the activity generated by a surface of matched

uniform chromaticity.

The edge-induced, filling-in phenomenon known as the Watercolor Effect (WCE) is ideal for distinguishing between these two hypotheses (Figure 2a). In the WCE, the color of the inner contour of a pair of distant, chromatic contours modifies the appearance of an interior region that is physically identical to the background in such a manner that it appears as a uniform, desaturated hue (Pinna, Brelstaff, and Spillmann 2001). The WCE propagates over too large a visual angle to be attributed to light spread in the eye. However, to date, investigation of the WCE has provided little evidence concerning the neuronal underpinnings of the phenomena. Importantly, based on theoretical and psychophysical observations (Devinck et al. 2014), the thin chromatic edge-transient that induces the WCE is expected to preferentially activate edge-sensitive double opponent cells in area V1 rather than surface responsive single opponent cells. While Coia et al. (2014) obtained evidence from the visual evoked potential (VEP) correlates with psychophysical responses for the WCE, insufficient resolution of the VEP make it impossible to localize the areas implicated in the processing. Finally, both the large extent of the filling-in and the sensitivity of the phenomenon to the curvature of the inducing contours (Gerardin et al. 2014) point to the involvement of cortical areas beyond V1 and V2.

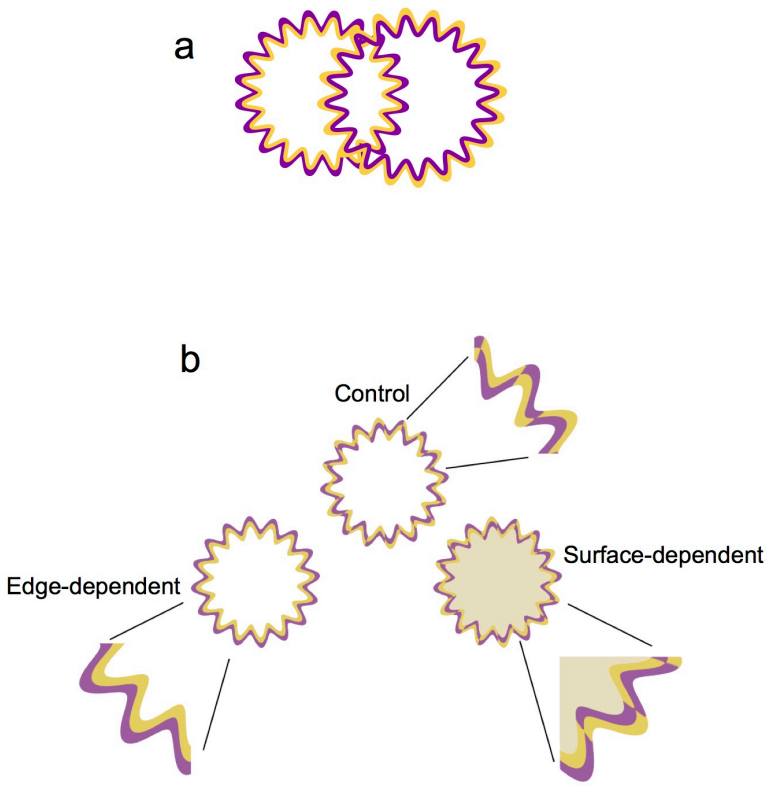

Figure 2: Watercolor effect demonstration and fMRI stimuli used in the main experiment.

a. Stimulus demonstrating the WCE. Two intersecting circular chromatic contours, each with its radius frequency modulated, thereby defining three regions. In the left region, the interior region takes on the hue of the bright orange interior contours showing the WCE while in the right, the dark purple contours generate no filling-in. The region of intersection that is enclosed by both types of contours is described by some observers as showing a weaker filling-in color. b. Examples of the three classes of stimuli used in the fMRI experiment. Beside each stimulus an enlarged detail of the contour is drawn to illustrate its structure Left: Edge-dependent stimulus inducing the perception of the WCE, as defined from (Gerardin et al. 2014); Center: The control stimulus is composed of braided contours and demonstrates a decrease in strength of WCE filling-in; Right: The surface-dependent stimulus was used to match the effect of the WCE filling-in, in which the interior chromaticity was defined from each subject's matches (from a preliminary paired comparison experiment inside the scanner). These examples are schematic in terms of the width and color of the contours and interior region, which are enhanced here for visibility of the stimulus structure used.

Here, we used functional magnetic resonance imaging (fMRI) combined with Multi-Voxel Pattern Analysis (MVPA) in order to localize those cortical areas associated with edge-induced, color filling-in of the WCE phenomenon and those responding to a matched uniform chromaticity. To validate the functional imaging findings, we compared the blood oxygen-level dependent (BOLD) signal with the psychophysically estimated strength of the WCE. The MVPA results demonstrate that neural activity supporting the WCE is distributed across multiple hierarchical levels and streams in the visual system, but only a small set of higher order dorsal areas are related to the perceptual strength of the phenomenon. Combining Dynamic Causal Modeling (DCM) with Bayesian Model Selection, shows that the areas most associated with the WCE exhibit a feedback modulation on area V1 and the ventral stream, and the directions of these modulations are reversed for the matched uniform chromaticity. The results suggest that the WCE color filling-in depends, at least in part, on a dorsal stream influence on ventral color areas. They support the idea that edge-induced filling-in and surface-dependent color are processed through distinct networks across the cortical hierarchy engaging ascending and descending pathways.

2. Material and Methods

2.1. Observers

Sixteen observers (10 female, mean±SD age: 28±4 years) participated in the study. All observers had normal color vision (Ishihara test and Panel D15), normal or corrected to normal vision and were all right-handed. The study was approved by the local ethics committee (ID-RCB 2012-A000123-40) and all participants gave written informed consent.

2.2. Stimuli

The stimuli were created with Matlab R2010b (Mathworks, MA., U.S.A.), and displayed with the PsychToolbox extensions (Brainard 1997, Pelli 1997). Spectral and luminance calibrations of the display were performed with a PR-650 SpectraScan Colorimeter (Photoresearch) and used for screen gamma-correction in stimulus specification. All stimuli were displayed on a white background (580 cd/m2, CIE xy = 0.29, 0.30). Stimuli in the main experiment were defined by a virtual circle (8 degree diameter, i.e., 4 degree eccentricity when centrally fixated) whose radius was modulated sinusoidally. A subset of observers was also tested in independent sessions with a 4 degree diameter stimulus as a comparison to conditions from previous psychophysical studies (Devinck et al. 2014, Gerardin et al. 2014). These measures served as a test for repeatability of the pattern of results. The contour pairs that defined stimuli in all conditions had a combined width of 16 min, i.e., 8 min each. The classical WCE was generated with a pair of continuous adjacent contours and will be referred to as the *test or edge-dependent* condition (Figure 2b, left). The outer contour appeared purple (CIE xy = 0.32, 0.19) and the inner orange (CIE xy = 0.48, 0.34). Control stimuli were defined by interleaving the interior and exterior contours in a braid (Figure 2b, center) (Devinck and Knoblauch 2012). All stimuli were also specified in the DKL color space (Derrington, Krauskopf, and Lennie 1984) with purple and orange contours at azimuths of 320 and 45 deg, respectively. Additional information is described in the *Stimuli and conditions* paragraph in the section fMRI Design and procedure: the WCE experiment.

2.3. Psychophysical Procedures

Prior to the collection of fMRI data, all observers performed the Maximum Likelihood Difference Scaling (MLDS) task, (Knoblauch and Maloney 2012, 2008, Maloney and Yang 2003) in order to measure the perceived magnitude of the WCE from the stimuli *in situ*, in the scanner, for each observer, following the procedure introduced by Devinck and Knoblauch (2012). To confirm that observers responded according to the fill-in color and not on the basis of other stimulus feature(s),

equal numbers of test and control stimuli were interleaved in the session. On each trial, a randomly selected triad of either test or control patterns was presented with three luminance elevations of the orange contour (a, b, c) chosen from a series of 10, with a < b < c. Stimulus b was always the upper stimulus in the middle, and stimuli a and c were randomly positioned below on the left or right side, respectively (Figure 3a). In each session, there was a random presentation of the 10!/(3! 7!) = 120 unique triads from the series of 10 luminances of the orange contour, equally-spaced in elevation from the equiluminant plane in DKL space from 0.1 to 0.9, for both test and control stimuli triads, thus yielding a total of 240 presentations. Frequency and amplitude of the contour were set to 20 cycles per revolution (cpr) and 0.2, respectively, with values shown previously to generate a strong WCE (Gerardin et al. 2014). On presentation of a triad, the observer was instructed to fixate each pattern until he/she could choose which of the two bottom patterns (left or right) was most similar to the upper pattern with respect to the color of its interior region. The observer's response initiated the next trial. No feedback was provided to observers. Difference scales for test and control stimuli were estimated from the session by maximum likelihood using the mlds function from the MLDS package (Knoblauch and Maloney 2008) in the open source software R (RCoreTeam 2015). The scales estimated are based on a signal detection model of the decision process and have the property that equal scale differences are perceptually equal, i.e., they are interval scales. When parameterized so that each response has unit variance, the scales can be interpreted in terms of the signal detection parameter *d'* (Knoblauch and Maloney 2012). Additional details on the procedure and modeling can be found elsewhere (Devinck et al. 2014, Devinck and Knoblauch 2012).

2.4. fMRI Design and procedure

Observers were presented sequences of stimuli selected from the three conditions described below. To control for attentional factors, during all functional MRI experiments, observers performed a task requiring them to detect a change in orientation of the fixation cross. Observers were instructed to press a button when they detected a change in the fixation cross from '+' to 'x'. Performance for this task was 79% ±3% (SD) correct. All stimuli were back-projected using a video-projector on a translucent screen positioned at the rear of the magnet. A subject, lying on his/her back in the scanner, viewed this screen via a mirror fixed on the head coil. Viewing distance was 222 cm.

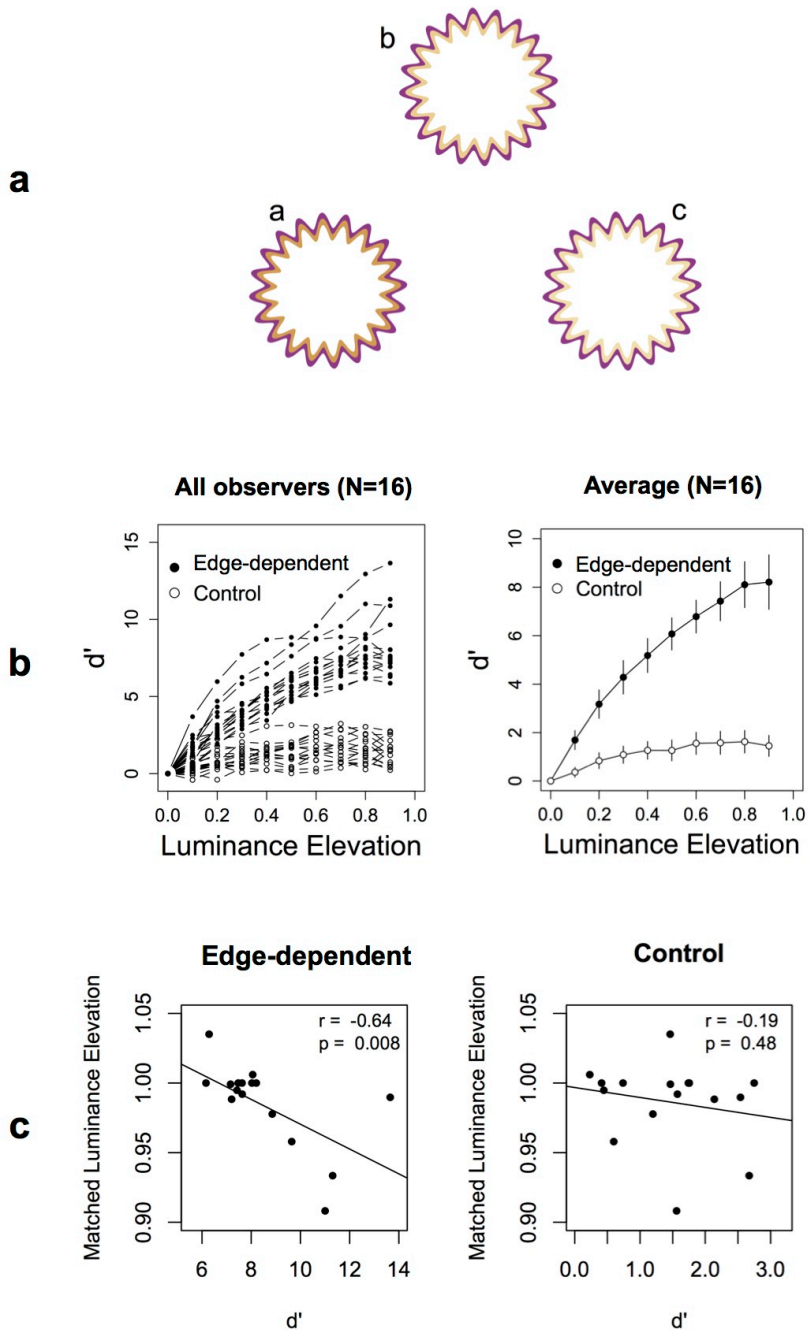

Figure 3: Psychophysics: stimulus configuration and estimated perceptual scales.

(a) Example of the triad configuration used in the MLDS task performed in the session before scanning. A triad consisted of either edge-dependent (WCE) or braided control stimuli with three unique ordered luminance elevations (a, b, c, selected randomly from 10 elevations in total) of the interior orange contour. Stimulus b was always the upper stimulus in the middle, and stimuli a and c were randomly positioned on the left or right. The observer had to fixate each pattern until he/she could choose which of the two bottom patterns (left or right) was most similar to the upper pattern with respect to the color of its interior region. (b) Left all response scales for the sixteen observers as a function of luminance elevation for edge-dependent stimuli (black) and control stimuli (white). The signal detection model used to fit the observer's data permits the ordinate values to be expressed in

terms of the signal detection measure d'. Right average response scales for the sixteen observers as a function of luminance elevation for edge-dependent and control stimuli with 95% confidence intervals. (c) Correlations of MLDS results with paired-comparison matches. Results for each observer of the luminance elevation that best matched the edge-dependent stimuli (WCE) against the peak MLDS response for edge-dependent (left) and for control stimuli (right).

2.4.1. Stimuli and conditions

Three conditions were used. *a) Edge-dependent stimuli*: Based on our previous studies (Gerardin et al. 2014) and the psychophysical experiments performed in the scanner, stimuli were chosen to generate strong WCE. Three contour frequencies (16, 18 and 20 cpr) were used, with two contour amplitudes (0.16 and 0.20) and with three luminances of the orange contour (elevations of 0.7, 0.8 and 0.9 in DKL space) (Figure 2b left). *b) Control stimuli:* Control stimuli were identical except that the contours were interlaced or braided and generated little filling-in (Figure 2b center). *c) Surface-dependent stimuli:* These stimuli had braided contours similar to the control, but the region interior to the contours was set to a uniform chromaticity and luminance that matched the interior, fill-in color of the edge-dependent stimuli based on a paired-comparison experiment performed inside the scanner for each subject (Figure 2b right). These stimuli were designed to generate the same appearance as the uniform interior of the WCE stimulus but with the bounding contour of the control stimulus.

For the paired-comparisons (performed in the scanner prior to scanning), observers were presented simultaneously with a test and a control stimulus (braided contour) for which the appearance of the interior was controlled by adjusting the luminance elevation (in DKL color space) of the interior with respect to the equiluminant plane at constant vector length, thereby continuously varying the colorimetric purity of the interior region. Observers judged which interior region of the pair was more orange and the match was assigned to the value at which the probability of a choice was estimated to be 50%.

2.4.2. Procedure

Each observer was tested in one session comprising six scanning runs. Each run consisted of 12 blocks plus 5 fixation intervals (including one at the start, one at the end, and one between each 3-block sequence) and lasting for 272 sec. Three conditions were tested: a) stimuli inducing WCE (edge-dependent), b) null filling-in stimuli (control) and c) surface-dependent stimuli. Each block

comprised 20 stimuli of one condition and was repeated four times. Blocks were presented in a randomized sequence. Each stimulus was presented for 500 ms followed by a blank interval (300 ms).

2.5. Mapping regions of interest

In a second session, for each observer, we identified: 1) retinotopic areas V1 and V2, 2) the human V4 (hV4, also identified by retinotopic mapping) and Lateral Occipital complex (LO) and 3) motion-related areas (V3B/KO, hMT+/V5). We functionally localized early visual areas V1 and V2, dorsal retinotopic areas V3, V3A and V7 and ventral retinotopic areas V3v and hV4 based on standard retinotopic mapping procedures (Warnking et al. 2002, DeYoe et al. 1996, Sereno et al. 1995, Engel et al. 1994). We also functionally localized LO (Kourtzi and Kanwisher 2001), V3B/KO (Dupont et al. 1997), and hMT+/V5 (Tootell et al. 1995) with a block related paradigm detailed below (Figure 4).

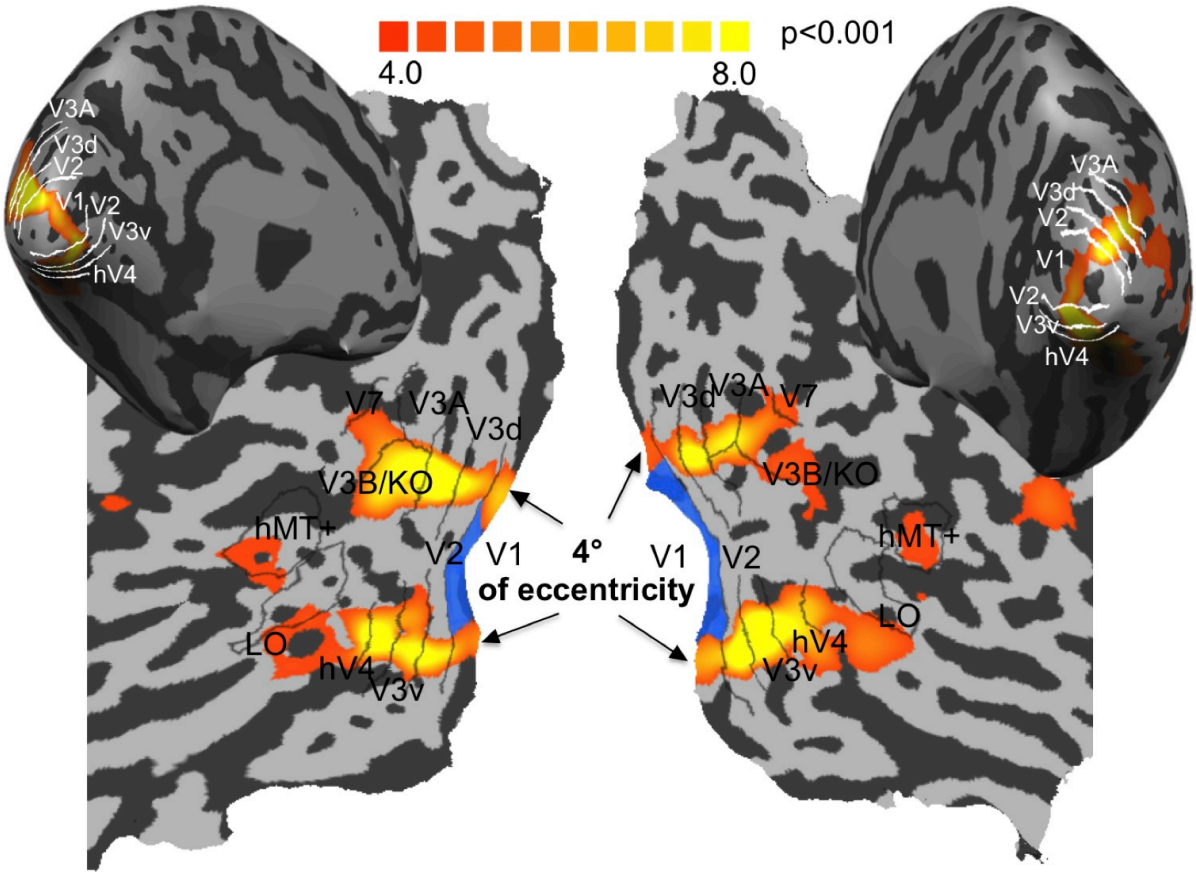

Figure 4: ROIs delineation.

Mapping of Regions of Interest (ROIs). We identified: 1) retinotopic areas, 2) the Lateral Occipital complex (LO) and 3) motion-related areas (V3B/KO, hMT+/V5).). Functional activation (from red to yellow, for low to high activation) from the braided control (4 deg eccentricity) versus fixation conditions was projected onto each inflated or flattened hemisphere (Left: left hemisphere; Right: right hemisphere for one subject). The blue shading in

area V1 indicates the retinotopic regions interior to the activation by the 4-degree eccentric contours of the control stimulus, based on the eccentricity maps obtained from the retinotopy mapping experiment. This defines the region interior to the stimuli from which the voxels were defined for the analyses in order to exclude voxels responding directly to the contour.

2.6. fMRI Data Acquisition

All experiments were conducted using a 3-Tesla Philips Achieva MRI scanner at the Grenoble MRI facility IRMaGE, France. In each session for each individual, a high-resolution T1-weighted structural image (3D TFE sequence, acquisition matrix 240x256x180, TR/TE: 25/2.3 ms, flip angle 9°, 1x1x1mm resolution) and series of T2*-weighted functional images (EPI MS-FFE, acquisition matrix 80x80, 30 slices, TR/TE: 2000/30 ms, flip angle 80°, acquisition voxel size 3x3x2.75mm, reconstructed voxel size 3x3x3mm) were collected with a 32 channel SENSE head coil.

In order to control for fixation stability, the position of the left eye over the course of all experiments was monitored with an ASL EyeTracker 6000. No systematic deviations from the fixation point were observed and no data were excluded from the analysis.

2.7. fMRI Data Analysis

The fMRI data were analyzed using Brain Voyager QX (Brain Innovations, Maastricht, The Netherlands). Preprocessing of the functional data consisted of slice-scan time correction, head movement correction, temporal high-pass filtering (2 cycles) and linear trend removal. Individual functional images were aligned to each corresponding anatomical image that was used for 3D cortex reconstruction, inflation and flattening. The volume time-course datasets were convolved with a canonical hemodynamic response function. No spatial smoothing was applied except for the group whole brain analysis (Gaussian filter; full-width at half maximum: 6 mm kernel) after normalization of all data to a common referential (Talairach space). Multi-Vovel Pattern Analyses (MVPA) were performed on each individual with Matlab R2010b (www.MathWorks.com) and SVM^{light} (Joachims 1999). Dynamic Causal Modelling (DCM) analyses were performed using Matlab (R2014a- with SPM12 (6906) with supplementary analyses using the open source software R (R Core Team, 2017).

2.7.1. Mapping Regions of Interest

For retinotopic areas we computed correlation analyses with a sinusoidal function with 16 lags to obtain power and phase maps for both the eccentricity and the polar mapping. Phase maps were thresholded at a correlation of 0.2 and projected on the cortical flat maps. The borders between visual areas were identified as phase reversals on the polar phase maps, with simultaneous visualization of the eccentricity map to ensure that the borders ran perpendicularly to the eccentricity gradient. We identified V1 and V2, dorsal areas up to V7 and ventral areas up to hV4. In addition, for area V1, a set of voxels was selected that was retinotopically interior to those significantly activated by the contours of the control stimulus (4-degree eccentricity, blue shading in Figure 4) for use in the subsequent MVPA analysis. This selection was based on the data from the retinotopic mapping experiments, i.e., eccentricity maps defined using expanding concentric rings. This allows excluding for classification patterns on the basis of voxels that respond in each condition to the contour *per se*.

Using the general linear model (GLM) including fixation periods and movement correction parameters as covariates of no interest, we identified LOC, hMT+/V5 and V3B/KO. LOC was defined as the set of contiguous voxels in the ventral occipito-temporal cortex that showed significantly stronger activation ($t(165)>4.0$, $p<0.001$ uncorrected) for intact compared to scrambled images of objects (Kourtzi and Kanwisher 2001). hMT+/V5 (Tootell et al. 1995) was defined as the set of contiguous voxels in the lateral occipito-temporal cortex that showed significantly stronger activation ($t(165)>4.0$, $p<0.001$ uncorrected) for moving than for static random dots. V3B/KO was defined as the set of contiguous voxels anterior to V3A that showed significantly stronger activation ($t(165)>4.0$, $p<0.001$ uncorrected) for random-dot displays that defined relative rather than transparent motion (Dupont et al. 1997).

2.7.2. Multi-voxel pattern analysis

Multi-voxel Pattern Analysis (Haynes and Rees 2005, Kamitani and Tong 2005), using linear Support Vector Machine (SVM) classifiers followed by cross-validation procedures, was used with Matlab R2010b (www.MathWorks.com) (Gerardin, Kourtzi, and Mamassian 2010) and SVM^{light} (Joachims 1999). For each Region of Interest (ROI: retinotopic areas, LO, V3B/KO and hMT+/V5), voxels were sorted according to their response (t-statistic maps based on the GLM) to all stimulus conditions compared to fixation baseline across all experimental runs. The number of voxels was selected across

individual ROIs and observers by restricting the pattern size to those voxels that showed a significant ($p<0.05$ uncorrected) t value. One hundred voxels from both hemispheres were included for each ROI and subject. For area V1, voxels were chosen to be at retinotopic eccentricities inferior to those activated directly by the control stimulus contour at 4 degrees (blue region of V1 in Figure 4). For other visual areas, it was not possible to exclude such voxels for all observers and maintain 100 voxels per ROI. Each voxel time course was normalized (z-score) separately for each experimental run in order to minimize baseline differences. For each subject and each condition, all time series data points of the corresponding experimental block were averaged in order to generate the data vectors for the multivariate analysis. Data vectors were selected according to the comparison of interest and split into a training sample comprising the data of five runs and a test sample comprising the remaining run. A six-fold cross-validation was performed leaving one run out (test sample). For each subject, accuracy rates were averaged (number of correctly assigned test patterns/total number of assignments) across the cross-validation runs. Statistical significance across conditions was evaluated using linear mixed-effects models with observer treated as a random effect (Pinheiro and Bates 2000). A recent study questioned the use of parametric significance tests with MVPA based on simulations that revealed skewness of the null-distribution (Jamalabadi et al. 2016). Diagnostic plots of the residuals from the mixed-effect model fits (residuals vs fitted values, and quantile-quantile plots, Figure S3), however, revealed no systematic deviations from model distribution or variance assumptions that would warrant discounting the model or preferring transformed values of the response variable.

2.7.3. Control analyses for the MVPA

Control analyses were obtained by shuffling to ensure that the classification accuracies observed were not due to random statistical regularities in the data. The same classifiers were run with randomly assigned category labels to the activation pattern. Accuracies were not significantly different from chance, supporting that the classification results could not be attributed to random patterns in the data.

2.7.4. Dynamic Causal Modeling for BOLD responses

Using Dynamic Causal Modeling (DCM), we explored changes in the effective connectivity, i.e. the influence exerted by one region on the others, and how information is propagated through these regions (see (Friston 2011) for details about the operational distinction between functional and

effective connectivity) in response to our protocol. DCM considers the brain as a deterministic system whose response in a region or part of a cortical network, is determined by inputs to that region. We defined causal models of connectivity between selected ROIs to make inferences about underlying mechanisms expressed in our BOLD data. These models included 1) intrinsic (or endogenous) connections that quantify the effective fixed connectivity between the model nodes, i.e. the changes in activity in a target region when activity in the source region changes (matrix A in DCM), 2) stimulus-related inputs that define how the model responds to task-related inputs (matrix C in DCM), and 3) possible input modulations that define how the effective connectivity is influenced by experimental factors (matrix B in DCM). Following model estimations, DCM provides a Bayesian method (Bayesian Model Selection, BMS) for selecting the best model of coupling that explains connectivity changes underlying task-related brain responses (Friston, Harrison, and Penny 2003). Note that effective connectivity between regions does not imply underlying direct anatomical connections (Friston 2011).

2.7.4.1. Model structure: Following BMS, each model is attributed a probability according to its power to describe the data. These probabilities sum to one, so the number of models in the model space must be limited to prevent excessive dilution of the maximum probability. One way to achieve this is to limit the number of areas considered. In our case, three regions were considered in each model definition because it is the lowest number necessary to translate our hypothesis concerning the connectivity modulations between dorsal and ventral streams. For example, region V1 (as receiving visual input), region V3A (that best classified edge-dependent stimuli) and region LO (that best classified surface-dependent stimuli). As early visual areas are strongly interconnected, we considered all possible connections between and within areas in our endogenous matrix. Our control stimuli entered all models as a driving input to V1.

2.7.4.2. Connectivity Modulation: We designed a model space in terms of which a subset of connections was modulated by the edge-dependent or the surface-dependent stimuli. We reasoned that both types of stimuli have a differential effect on endogenous connections of the network for each subject. We did not consider bidirectional modulations because we were only interested in the strongest direction. V1 was always connected to the two other areas in all models, but we included the hypothesis that there was no differential modulation between them. Because of the importance of

V1 intrinsic horizontal connections for contour integration (Stettler et al. 2002), for each model we included a self-modulation of V1. This resulted in the model space described in Figures 6a and S2a with 12 possible modulations.

2.7.4.3. Model comparison: All twelve models were applied successively with the functional datasets from each subject and each condition (edge-induced or surface-dependent). They entered a Bayesian model selection procedure that identified which of the competing models best predicts each dataset. For each model, the model evidence, i.e. the probability of observing the measured data given a specific model, was computed based on the free energy approximation. Model evidence was used by the Bayesian model selection procedure to rank the models. The protected exceedance probability defined the relative superiority that a model was more likely that any other model considered, given the group data (Rigoux et al. 2014). For group-level Bayesian model selection we then considered a Random-effects (RFX) analysis to account for between-subject variability (Stephan et al. 2009).

3. Results

3.1. Estimated perceived strength of WCE

Perceptual scales were estimated by Maximum Likelihood Difference Scaling (MLDS), method of triads (Knoblauch and Maloney 2008, Maloney and Yang 2003), a powerful method for identification of perceptual correlates of BOLD responses (Bellot et al. 2016, Yang, Szeverenyi, and Ts'o 2008). A sample trial illustrating the stimulus configuration is shown in Figure 3a. The approach is based on the idea that when the perceived filling-in of the upper stimulus is judged equally often to be as similar to the lower left as to the lower right stimulus, then the perceived differences between a and b and between b and c are equal. This allows scale values to be estimated by maximum likelihood based on a signal detection model of the observer decision rule (Knoblauch and Maloney 2012, Maloney and Yang 2003). Individual (left panel) and average (right panel) perceptual scales estimated by MLDS (Figure 3b) display the dependence of the strength of the filling-in on the luminance of the orange contour for test (filled circles) and control (open circles) stimuli. With test stimuli the filling-in response increases with the elevation of the luminance of the interior contour with respect to an equiluminant plane in DKL space, while for the control stimuli responses are attenuated or absent (Devinck et al. 2014, Devinck and Knoblauch 2012). The differences in response magnitude between edge-

dependent and braided control stimuli confirms that observers respond to the fill-in color and not the contours, and reveals significant differences in the strength of the WCE across observers. Figure 3c shows for all observers two scatterplots of the luminance elevation that best match the WCE against the peak MLDS response for the edge-dependent stimulus (left panel) and the control stimulus (right panel). As the luminance elevation of the matched stimulus decreases in DKL space, colorimetric purity increases. Thus, lower luminance matches correspond to a higher matched colorimetric purity and therefore, a stronger perceived fill-in color in the edge-dependent stimulus. The strength of the WCE is significantly correlated with the matched color value ($r=-0.62$, $p=0.01$) for the edge-dependent but not for the control stimulus ($r = -0.19$, $p = 0.48$). Thus, individual differences in the strength of the WCE covary similarly for both the scaling and matching data.

3.2. Rationale of test and control conditions for imaging experiments

Table 1 shows the pattern of feature differences among the three stimulus conditions that we used and on the basis of which we can make comparisons and inferences about the role of particular Regions of Interest (ROIs) on edge and surface dependent color appearance. Because each comparison varies for two features, no comparison uniquely defines the information treated by a ROI, instead the interpretation of the source of the activation is constrained by the combined classification across all three comparisons.

Table 1: Stimulus and response differences tested by classifiers.

	Edge continuity	Interior chromaticity	Interior perceived color
Edge-dependent vs Control	**X**		**X**
Surface vs Control		**X**	**X**
Surface vs Edge-dependent	**X**	**X**	

The X's indicate the features on which the comparisons differed. The pattern of responses across the three classifiers and not a single classification constrains the specificity of the region tested.

For example, significant differences for the edge-dependent vs. control comparison could result from either the local difference in edge continuity of the inducer or from the difference in perceived color of the interior region. However, if the same ROI only weakly differentiates the surface-dependent vs. edge-dependent, then the influence of the edge continuity on the response can be excluded since

these stimuli also differ with respect to that feature. In this case, we infer that the area's response is implicated in the treatment of filling-in. If in addition the ROI weakly differentiates surface vs. control conditions, then it implies that the ROI's response does not depend strongly on the surface chromaticity since that feature is a common feature to the two comparisons showing weak classification. On the other hand, if an ROI strongly responds to both the edge-dependent vs. surface and the surface vs. control conditions but not the edge-dependent vs. control, then we can conclude that it responds on the basis of a purely stimulus bound feature, the interior chromaticity difference, since that is the only common feature.

3.3. BOLD activation related to the WCE

A whole-brain analysis using a General Linear Model (GLM) was undertaken so as to detect significant BOLD activation related to the WCE. With this aim, stimulus blocks from each of the three conditions were contrasted with the other two in a group analysis. No significant activation was revealed ($p>0.05$, Bonferroni corrected for all brain voxels).

3.4. Discriminating edge-dependent and surface-dependent processing with MVPA

To obtain a more fine-grained analysis of the contrasts, MVPA was applied to ROIs identified by functional localizers. We assumed that if a ROI is involved in edge- or surface-dependent color appearance, contrasts in the stimuli would generate a significantly higher accuracy than chance according to the logic expressed above and with respect to Table 1. Signal-to-noise ratios tested by a linear mixed-effects model across cortical areas and with a random observer intercept showed no significant differences that would bias such comparisons (Likelihood ratio test: $\chi^2(9) = 7.4; p = 0.6$, Figure S1).

MVPA was used to test how activity patterns of voxels in retinotopic areas (V1, V2), dorsal areas (V3d, V3A, V7, V3B/KO, hMT+/V5) and ventral areas (V3v, hV4, LO), are sensitive to the WCE. As demonstrated previously (Devinck et al. 2014, Gerardin et al. 2014) and confirmed in the psychophysical experiments above, WCE filling-in is absent or strongly attenuated when the stimulus contour is braided. Classifiers were run to identify regions that respond preferentially to these particular properties of the edge-dependent stimuli (WCE). Analyses were also conducted to reveal visual regions that respond to the processing of the surface-dependent stimuli matched in color

appearance to the WCE. For each area, three classifiers tested specific hypotheses: (I) that accuracy was higher for the classification of edge-dependent than control stimuli; (II) that accuracy was higher for the classification of surface-dependent than control stimuli and (III) that accuracy was higher for the classification of edge- than. the surface-dependent color stimuli. All results reported below were corrected for multiple testing (Holm 1979).

The 95% confidence intervals of the classification accuracies for all of the non-shuffled conditions (Figure 5, left bar in each of the three pairs in each graph and Table S1) exclude chance level at 50% while those for the shuffled control analyses (Figure 5, right bar in each pair) all include chance. These results suggest that processing of the stimuli tested is distributed across many areas. However, not all areas responded differentially to the three classifiers. For 4 areas (V2, V3d, V7 and hMT+/V5), the classification accuracies showed no selectivity for the conditions (Table S2, linear mixed effects model: V2: $F(2, 30) = 2.96$, $p = 0.07$; V3d: $F(2, 30) = 0.87$ $p = 0.42$; V7: $F(2, 30) = 1.53$, $p = 0.23$; hMT+/V5: $F(2, 30) = 0.28$, $p = 0.76$), and except for area V2, which classified robustly for all three MVPA rules, displayed relatively low prediction accuracies. Significant differences between the MVPA rules were observed in V1 (Table S2, linear mixed-effects model: $F(2,30)=16.58$, $p<0.001$) and in ventral stream areas (V3v: $F(2,30)=7.61$, $p<0.01$); hV4: $F(2,30)=11.78$, $p<0.001$; LO: $F(2,30)=17.14$, $p<0.0001$). In dorsal areas, differences between the MVPA rules only attained significance for areas V3A ($F(2,30)=47.05$, $p<0.00001$) and V3B/KO ($F(2,30)=27.27$, $p<0.00001$) (Table S2).

In independent sessions, the measures were repeated in three of the observers using stimuli of 4 deg diameter. The average of their results are indicated by the yellow points. In general, they replicate the same patterns as for the larger stimuli and support the robustness of the findings.

Ventral and dorsal stream areas systematically differ according to the patterns of classification accuracy across the three MVPA rules. Ventral areas (V3v, hV4 and LO) displayed higher accuracies for the surface-dependent vs. edge-dependent classifier that additionally were significantly greater than for the edge-dependent vs. control rule (Table 2, Tukey test, V3v: $t(885) = -4.3$, $p < 0.001$; hV4: $t(885) = -2.5$, $p < 0.05$; LO: $t(885) = -5.2$, $p < 0.001$). If the classifiers were driven by the edge continuity, then both rules would be expected to indicate high prediction accuracy. Instead, the results suggest that the classifications depended on the interior chromaticity of the stimuli, an interpretation reinforced by the high accuracy of the surface-dependent vs. control comparison (significantly greater

than the edge-dependent vs. control condition for areas V3v and hV4: Table 2, Tukey test, I vs II: V3v: t(885) = -3.1, p < 0.01 and hV4: t(885) = -4.7, p < 0.001) that differed in the interior chromaticity.

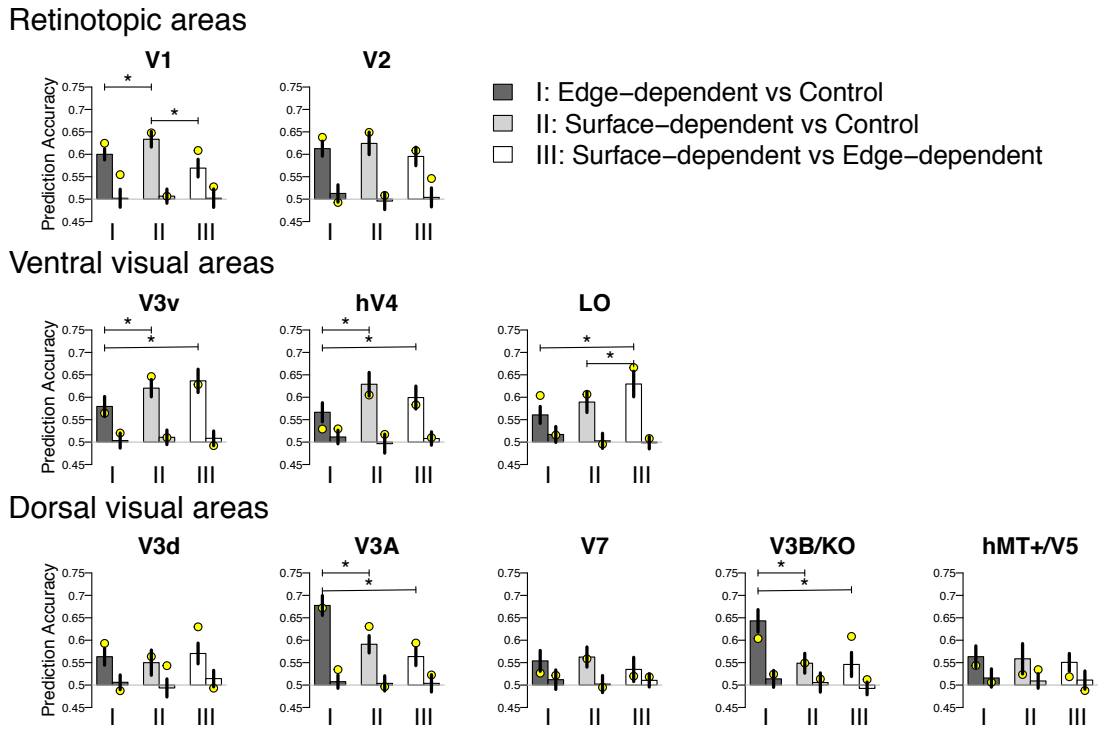

Figure 5: Multi-Voxel Pattern Analysis (MVPA) from fMRI data.

Classification accuracies across the 10 regions of interest for the three classifiers: (I) Edge-dependent vs. control (dark grey); (II) Surface-dependent vs. control (grey); (III) Surface-dependent vs. edge-dependent (white). Mean classification accuracy is based on 100 voxels per area in both hemispheres. Each bar plot shows results for 3 pairs, unshuffled (left) and shuffled (right), of classifiers. Error bars indicate 95% confidence intervals. Significant differences (p < 0.05) controlled for multiple comparisons (Tukey test) indicated by horizontal bars and star. The yellow points indicate average results of a repeated experiment on three of the observers using a 4 deg stimulus.

Dorsal stream areas (V3A and V3B/KO) showed a complementary pattern of results in which classification accuracy was highest for the edge-dependent vs. control condition and significantly higher than for the other two rules (Table 2, Tukey test, V3A, I vs. II: t(885) = 6.50; p < 0.001; I vs. III; t(885) = 8.50, p < 0.001; V3B/KO: I vs. II: t(885) = 7.10, p < 0.001; I vs. III: t(885) = 7.30; p < 0.001). The high accuracy in these areas for the edge-dependent vs. control rule supports either a classification on the basis of the filling-in or the edge continuity. The surface-dependent vs. edge-dependent comparison, however, also differed in edge continuity so that the significantly lower response in this case argues in favor of filling-in and against a role for edge continuity. In addition, the

significantly lower response of the surface-dependent vs. control comparison argues against an important role for surface chromaticity *per se* in the classification behavior of these areas.

Table 2: Post-hoc t-tests (Tukey Honest Significant Differences test) for differences of classification accuracy between rules within areas.

MVPA rule	I - II		I - III		II - III	
	t	p	*t*	p	*t*	p
Retinotopic areas						
V1	-2.50	**3.3e-02**	2.30	5.6e-02	4.80	**5.2e-06**
V2	-0.88	6.6e-01	1.30	3.9e-01	2.20	7.6e-02
Ventral visual areas						
V3v	-3.10	**6.6e-03**	-4.30	**6.4e-05**	2.20	6.9e-02
hV4	-4.70	**1.0e-05**	-2.50	**3.8e-02**	-2.30	2.2e-01
LO	-2.10	8.2e-02	-5.20	**9.3e-07**	-3.00	**7.5e-03**
Dorsal visual areas						
V3d	0.99	5.8e-01	-0.54	8.5e-01	-1.50	2.8e-01
V3A	6.50	**4.7e-10**	8.50	**4.0e-17**	2.10	1.0e-01
V7	-0.64	8.0e-01	1.40	3.3e-01	2.10	9.8e-02
V3B/KO	7.10	**9.5e-12**	7.30	**2.2e-12**	0.20	9.8e-01
hMT+/V5	0.35	9.3e-01	0.94	6.1e-01	0.59	8.2e-01

Roman numerals refer to classification rules: (I) Edge-dependent vs. control; (II) Surface-dependent vs. control; (III) Surface-dependent vs. edge-dependent. Each column indicates the t- and p-values for the difference of a pair of classification rules. Significance levels for $p < 0.05$ are indicated in bold.

Areas V1 and V2 showed a similar pattern of prediction accuracy for the three classifiers, with higher response on the first two, edge-dependent vs. control and surface-dependent vs. control, but only in area V1 do the differences between comparisons attain significance. This pattern is consistent with processing of the fill-in color and the surface color since the responses are strongest whether the edge or the surface is the source of the perceived color and weakest when this attribute is matched. However, the strong activity for all three classifiers could indicate the presence of subpopulations that respond differentially to all of the dimensions along which these stimuli differ.

3.5. Relationship between behavior and neural processing of the WCE

The MVPA analysis reveals candidate areas implicated in the processing of the WCE filling-in, but does not necessarily indicate any relation to perception. To explore such a link, the Pearson product-moment correlations were calculated between the individual classification accuracies for the edge-dependent vs. control MVPA rule and the peak values for the estimated perceptual scales (d' from the MLDS task). None of the other rules should be related to the perceptual strength of the WCE and with one exception (area hMT+/V5, edge-dependent vs. surface: $r = 0.52$, $p = 0.04$), no significant positive correlations were observed. Figure 6 shows the scatterplots of classification accuracy vs. d' with regression lines for all ROIs with the correlations and significance levels under the hypothesis of no correlation indicated in each figure. Of all the cortical areas studied, only areas V3A and V3B/KO showed a significant positive correlation (V3A: $r = 0.66$, $p = 0.006$, 95% CI (0.237, 0.869); V3B/KO: $r = 0.5$, $p = 0.049$, 95% CI (0.004, 0.794)).

The correlation for area V3v is near the criterion of significance used throughout (V3v: $r = 0.47$, $p = 0.064$, 95% CI (-0.030, 0.785), and one can legitimately question whether its value differs significantly from that observed for area V3B/KO (Nieuwenhuis, Forstmann, and Wagenmakers 2011). An analysis of covariance to predict classification accuracy with respect to the covariate d' and the factor Area indicated a significant interaction ($F(9, 140) = 2.21$, $p = 0.025$), demonstrating the presence of at least one slope that differed significantly from 0. Inspection of the estimated coefficients for each area revealed significant slopes for areas V3A ($t(140) = 2.81$, $p = 0.006$), V3B/KO ($t(140) = 2.43$, $p = 0.02$) and V3v ($t(140) = 2.04$, $p = 0.04$). Post-hoc tests, however, showed that the slopes of both areas V3A and V3B/KO differed significantly from that of V3v (Tukey Honest Significant Differences test: V3v - V3A: $t(140) = -7.4$, $p < 0.0001$; V3v - V3B/KO: $t(140) = -4.81$, $p = 0.0002$) while the slope difference between areas V3A and V3B/KO is at chance (V3A – V3B/KO: $t(140) = 2.59$, $p = 0.2305$).

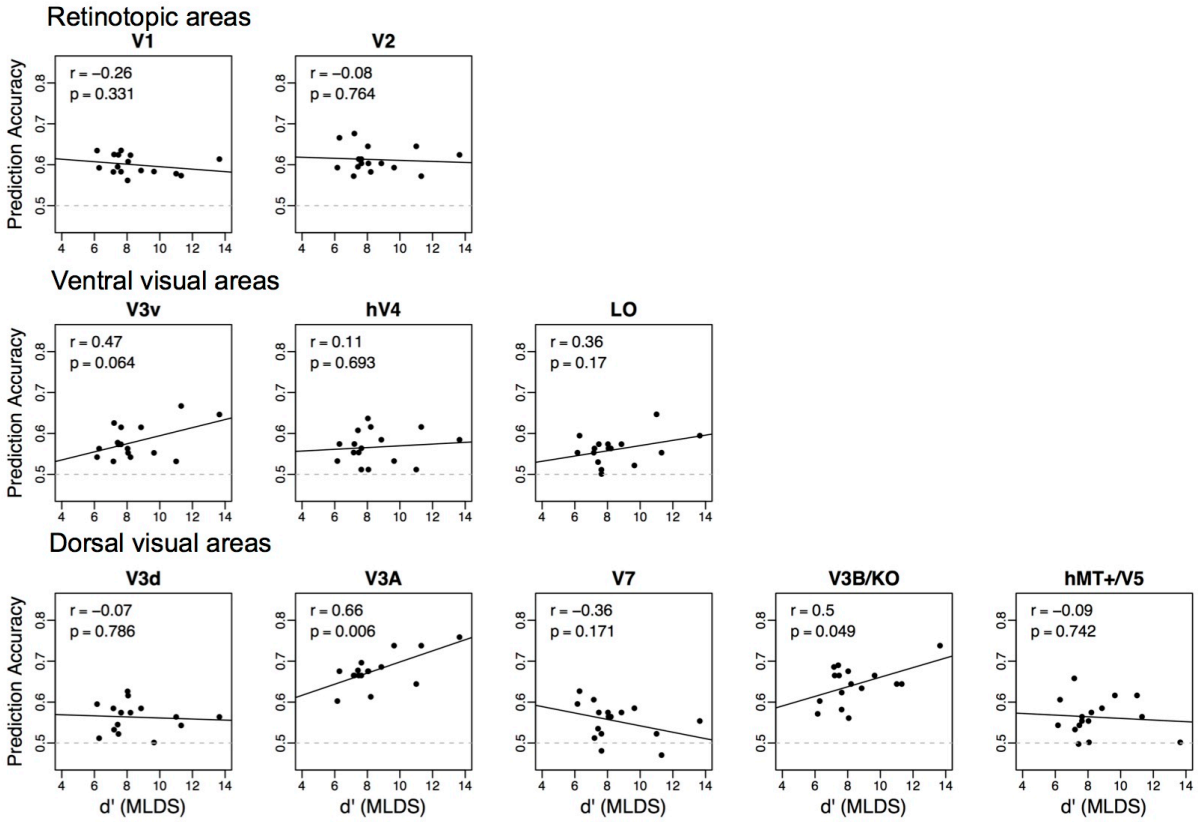

Figure 6: Correlating pattern classification and appearance.

Scatter plots of classification accuracies and peak MLDS scale values for all subjects (N=16) for the rule edge-dependent (WCE) vs. control stimulus for all ROIs with the best-fitting linear regression (solid line). Pearson product-moment correlations and p-values for the hypothesis that the correlation is 0 are indicated within each graph.

3.6. Effective connectivity for edge-induced and surface-dependent colors

The above results suggest a role for dorsal areas, particularly area V3A, in the WCE color filling-in. This is surprising because this area, and more generally, dorsal stream areas are little implicated in color processing. We hypothesized that such a role could be mediated through the action of these dorsal stream areas on the ventral stream, which has more typically been implicated in color processing. To evaluate this hypothesis, we used Dynamic Causal Modeling (DCM, (Stephan et al. 2010, Friston, Harrison, and Penny 2003)) and Bayesian Model Selection (BMS) to test the context dependent modulation among areas that showed differential responses to WCE and the uniform added chromaticity. To increase the power of the analysis, we limited evaluation to a triple of areas that included area V1, which represents the visual input at the cortical level, and two areas, V3A and LO, which based on the MVPA analyses, displayed the strongest contrast in their classification profiles

with respect to edge-dependent and surface-dependent stimuli. We hypothesized that these areas would be the most likely set of areas to differentiate between the two conditions. We considered models that contrasted the directional modulation of effective connectivity between each pair of the triple when compared with the braided control stimulus. We only tested models with unidirectional modulation of connections between each pair of areas to further reduce the number of models and because we were only interested in the strongest modulation of effective connectivity between area pairs. We considered a space of 12 models (4 patterns of modulation of connectivity of area V1 to V3A and LO times 3 possible relations between V3A and LO), representing plausible effective connectivities between these two areas and V1 (Figure 7a). V1 was always connected to the two areas, but we also included the hypothesis that there was no modulation between V3A and LO.

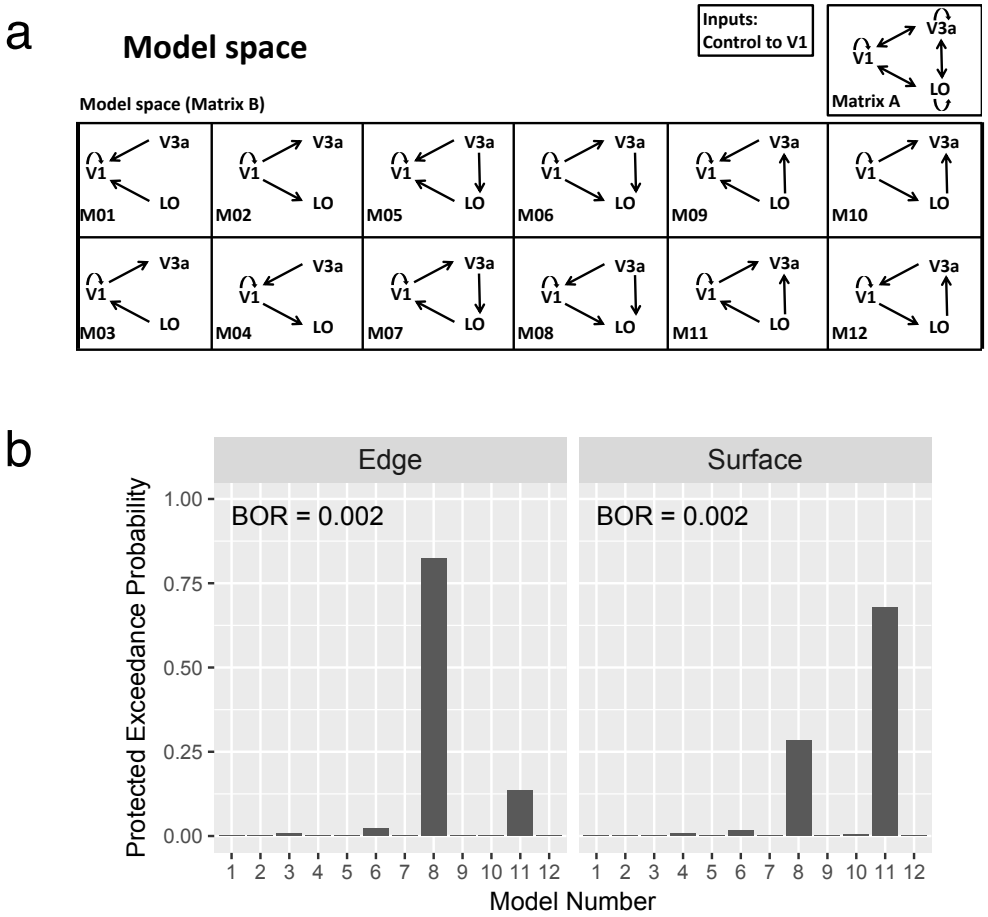

Figure 7. DCM Model space and protected exceedance probabilities for areas V1-V3A-LO.

a) Twelve models (M0-M12) of the modulations of effective connectivities between areas V1, V3A and LO were evaluated in the DCM analysis. The intrinsic connectivity (Matrix A) was taken to be complete within and between areas. Activations in response to the Control condition (braided stimuli, no perceived color) were used as input in V1, while activations in response to the Edge dependent or Surface dependent conditions could modulate

effective connectivity (Matrix B) according to each model. Each model included self-modulation in V1 and unilateral modulations between V1 and area V3A and between V1 and area LO, yielding 4 possible combinations. Additionally there could either be no modulation between area V3A and area LO (M01 to M04), unilateral modulation from area V3A to area LO (M05 to M08) or unilateral modulation from area LO to area V3A (M09 to M12). b) Protected exceedance probabilities (PEP) for each of the 12 models of modulation of the effective connectivities between the triple of areas indicated in (a) for the edge dependent (left) and surface dependent (right) stimuli. The Bayesian Omnibus Risk (BOR) is indicated in the upper left corner of each plot.

The protected exceedance probabilities (PEP) for the 12 models for each of the two conditions are shown in the bar graphs of Figure 7b for the edge- (left panel) and surface-dependent (right panel) conditions. The Bayesian Omnibus Risk factors (BOR) for the PEP profiles rejected the hypothesis that the models are equiprobable, thus, indicating that at least one differed significantly from the others (BOR = 0.002 for both cases). For the number of observers in our study, PEP values above 0.5 (the disambiguation threshold) are considered strong evidence in favor of the model (Rigoux et al. 2014). Thus, the edge-dependent condition strongly supported model M08 (PEP = 0.82) in which feedback connections from area V3A influence areas V1 and LO, and a feedforward connection from V1 modulates LO. The surface-dependent condition, however, strongly supported model M11 (PEP = 0.68), which displays a complementary set of modulations, i.e., LO exerting a feedback modulation on V1 and V3A and V1 a feedforward modulation of area V3A.

In the analyses above, areas V3A and LO were selected based on the clear contrast in their stimulus classification profiles. Concomitantly, we should expect that pairs of areas for which the classification accuracy was low and for which the profiles did not distinguish among the stimuli would not show evidence supporting different models of effective connectivity for the edge- and surface-dependent conditions. As a control, we examined a model space for the triple V1-V3d-V7 (Figure S2a). Both V3d and V7 showed low classification accuracies ($p < 0.6$) and there were no significant differences among the three comparisons performed (Figure 5). In addition, for all three of these areas, the correlations between the prediction accuracy and appearance measures were not significant (Figure 6). The analysis rejected the hypothesis that models were equiprobable (BOR < 0.05 for both edge and surface conditions), but the model for which the evidence was strongest was the same (M11) for both conditions (Edge: PEP = 0.039; Surface: PEP = 1.4e-4) (Figure S2b).

4. Discussion

In the current study, we compared cortical activity related to color appearance generated by either edge-induced filling-in or a uniform surface chromaticity and found them to be associated with complementary patterns of activity in dorsal and ventral visual streams. Three lines of evidence support these results. First, the MVPA technique showed a significantly different profile of classification performance for the edge-dependent vs surface dependent conditions in dorsal areas V3A and V3B/KO compared to ventral areas V3v, hV4 and LO. It is unlikely that differential deployment of attentional mechanisms could explain these results as attention was uniformly controlled in all fMRI experiments. Second, individual differences in the psychophysically measured strength of the WCE were significantly correlated only with the classification performance of the edge-dependent stimulus in dorsal areas V3A and V3B/KO. Finally, a DCM analysis showed that a model for which the data provided the strongest evidence was one in which the effective connectivity from dorsal area V3A modulated V1 and ventral area LO when the edge-dependent stimulus was compared with the control. In contrast, complementary modulations of LO onto V1 and V3A were found for the surface-dependent comparison with the control. Importantly, given that the whole brain GLM analyses failed to detect significant differences among the stimulus conditions, the results obtained depended critically on the sensitivity of the fine-grained MVPA (Sapountzis et al. 2010) and the DCM methods (Stephan et al. 2010, Friston, Harrison, and Penny 2003). Taken together, the results demonstrate that signals associated with edge- and surface-dependent information are not represented identically across visuals areas. The results support the hypothesis that information about surface-dependent and edge-induced colors is transmitted through distinct neural pathways.

The results do not rule out the participation of other areas in mediating edge- and surface-dependent color appearance. Indeed, the MVPA results reveal classification performance above chance in all visual areas, although they do not all significantly distinguish between the three conditions. For example, it would be interesting to extend the DCM analyses to include area V2 because of its implication in color processing (Roe and Ts'o 1999) and contour extraction (Lee and Nguyen 2001). This would entail examining a space of effective connections between four areas with an accompanying combinatorial increase in the number of possible models that would need to be tested. The larger number of models would dilute the power of such an analysis without, for example,

evidence from the MVPA to constrain the analysis to a relevant subset of hypotheses.

4.1. Implication of cortical areas in edge- and surface dependent color processing. While area V3A has not typically been associated with color processing, activity dependent markers revealed a sparse population of cells responsive to chromatic stimuli (Tootell et al. 2004). In addition, Hadjikhani et al. (1998) included V3A in a set of areas that generated greater responses to color than luminance modulated stimuli. More recently, Castaldi, et al. (2013), reported that V3A responds to chromatic spatial features. A source of its influence on color areas could arise from the recently described Vertical Occipital Fasciculus (Takemura et al. 2015) that links ventral and dorsal stream areas.

Area V3A exhibits response selectivity to stimulus features that are characteristic of the WCE. V3A has been shown to be selective for contour curvature (Caplovitz and Tse 2007), a characteristic of the inducer contours that is necessary to enhance the WCE but not sufficient to generate it alone, e.g., as shown by the braided control contours here and in previous work (Devinck et al. 2014, Gerardin et al. 2014). Area V3A is also implicated in filling-in phenomena, in that it was reported to show increases in activity during Troxler fading (Mendola et al. 2006). Areas V3 and V4 have recently been implicated in form-contingent filling-in induced by closed contours (Hong and Tong 2017). Similarly in macaque, area V3 has been implicated in filling-in phenomena (De Weerd et al. 1995). V3A also shows much larger spatial summation than do nearby cortical areas (Press et al. 2001) as might be expected for an area involved in filling-in over extended regions of the visual field. In fact, the large areas of the visual field over which the WCE can extend require mechanisms that are effective in the peripheral visual field. The implication of dorsal areas in large scale filling-in is consistent with the bias of peripheral visual field projections in the dorsal stream (Kravitz et al. 2013).

Early studies that reported retinotopically distributed neural activity due to brightness and color filling-in in area V1 (Sasaki and Watanabe 2004) have been disputed (Zweig et al. 2015, Cornelissen et al. 2006). Using fMRI, Cornelissen et al. (2006) found that surround-induced responses in V1 did not depend on surround intensity but instead could be attributed to an extended edge-response. This is consistent with voltage-sensitive dye-imaging in macaque that shows only edge responses to uniform fields of color and luminance in V1 (Zweig et al. 2015). Instead, V1 responses to contours have been attributed to figural effects rather than brightness or color filling-in (Kok and de Lange 2014,

Cornelissen et al. 2006, Lee and Nguyen 2001). Interestingly, the WCE stimulus is reported to demonstrate both a long-range color filling-in and a strong influence on figure-ground segregation (Von der Heydt and Pierson 2006, Pinna, Brelstaff, and Spillmann 2001). However, the results from area V1 are not likely to be explained uniquely by figure-ground processing. The classification results from this area (Figure 5) were based only on the voxels interior to those activated by the retinotopic positions of the contours (blue regions in Figure 4). The highest prediction accuracy for V1 significantly differed from that for both of the other classification rules and pitted the surface-dependent stimulus against the braided control, i.e., stimuli having the same braided contour. Instead, this supports that classification depended on the interior chromaticity and/or the perceived color and not local or long-range contour detection (Table 1). The poor correlation of the classification accuracies from this area with the perceived strength of the filling-in color (Figure 6) would support the former of these two interpretations.

Areas V3v, hV4 and LO displayed their highest classification accuracy to the surface chromaticity of the stimulus and not to the edge-induced fill-in conditions. In line with this outcome, the classification accuracies for these areas were not correlated with the strength of the perceived filling-in color. These results are consistent with previous studies that have demonstrated strong responses of these ventral stream areas to uniform surface colors (Brouwer and Heeger 2009, Parkes et al. 2009, Bouvier, Cardinal, and Engel 2008, Bartels and Zeki 2000, Hadjikhani et al. 1998, Sakai et al. 1995).

Area V2 presents a more complex pattern of results in that the classifiers trained for this area showed high accuracy for all three conditions with no selectivity between comparisons. This pattern of results is consistent with neurophysiological studies in primates that show diverse cell classes in area V2 including both those responsive to uniform color fields (Peng and Van Essen 2005, Roe and Ts'o 1999) and others that might be involved in contour detection and color induction (Roe and Ts'o 1999).

4.2. Role of double- and single-opponent cells. The initial segregation of responses to uniform color fields and edges in neural sub-populations occurs as early as area V1 and may reflect the differential responses of single- and double-opponent cells. The differing involvement of the dorsal and

ventral streams in edge and surface color responses could mean that these two cell classes project differentially to extra-striate cortical areas with the ventral stream areas being dominated by signals from single-opponent cells and the dorsal areas receiving projections predominantly from double-opponent cells.

Population responses to uniform color fields in area V1 are dominated by edge responses as demonstrated by optical imaging with voltage sensitive dyes (Zweig et al. 2015). This result could reflect that double-opponent cells are the most numerous color sensitive cell class in area V1 at the eccentricity examined. The less numerous single-opponent cells would be expected to respond as well to uniform fields. Thus, the typical cortical response signature in area V1 to a uniform color field could be expected to be a pattern of activity across cells with both types of receptive field profile. Isolated chromatic edges and edge transients, however, would be expected to generate a response profile dominated by double opponent cells.

Why then do chromatic edge-transients, such as in the case of the WCE, suffice to generate the appearance of a weakly colored field, filling-in the interior of a uniform region interior to the chromatic contour? It is significant that the color filling-in requires continuity of the bi-chromatic contour as shown here and elsewhere (Devinck and Knoblauch 2012) by the lack of effect for the braided control stimulus. This suggests a mechanism by which the coherence along the contour is integrated across a population of double-opponent cells' responses that we hypothesize would be mediated via paths from V1 to V3A. Such an interpretation is supported by results from adaptation studies demonstrating the role of contours in generating filling-in phenomenon (Coia and Crognale 2017, Hazenberg and van Lier 2013, van Lier, Vergeer, and Anstis 2009). We suppose then that the coherent contour would signal a uniform color field rather than just an edge-transient through the concurrent modulation of ventral color areas and V1 via dorsal stream areas implicated in contour integration.

4.3. Distributed nature of the responses. We find responses to the stimuli distributed across multiple streams and at multiple levels of the visual hierarchy, in line with an interareal network which is observed to be much denser than previously thought (Markov et al. 2013) and with evidence of the highly distributed nature of object processing (Konen and Kastner 2008). In area V1, the induced activity of voxels remote from the contours is unlikely to be mediated by monosynaptic lateral connections because intrinsic connections extend over 2 mm or less (Markov et al. 2011), and given a

cortical magnification factor of 6-13 mm/deg (Van Essen, Newsome, and Maunsell 1984, Daniel and Whitteridge 1961), it is unlikely that lateral activity spreads much larger than a degree (Angelucci and Bressloff 2006). This suggests that feedback projections from higher order areas play a role in the activation of area V1. The involvement of descending pathways is highly relevant given the role of feedback pathways in contextual processing (Zipser, Lamme, and Schiller 1996). Here, feedback projections from areas V3A and V3B/KO could be particularly significant given that we show that these areas code the behavioral response and that the dynamic causal model that best supports the data indicates that the effective connectivity from V3A to V1 is modulated differentially by the WCE. Recent human studies suggest that area V3A is relatively low in the cortical hierarchy (Michalareas et al. 2016) so that these findings are in line with previous evidence of activity at early hierarchical stages reflecting the content of consciousness (Leopold and Logothetis 1996). In the framework of predictive coding, that postulates that prediction errors are propagated up the cortical hierarchy and predictions down the hierarchy (Friston and Kiebel 2009), the descending prediction signal from area V3A might not be related to consciousness content but instead could be an aspect of figure-ground distinction which is thought to be represented in area V1 (Kok and de Lange 2014, Cornelissen et al. 2006, Von der Heydt and Pierson 2006) or alternatively a signal for a prior indicating the presence of a uniform color field.

5. Conclusion. Our results provide evidence that the visual system employs separate networks for the processing of surface colors that are generated by a field of uniform chromaticity and those due to filling-in induced by distant chromatic edges, so that the neural responses to a uniform and a filled-in color are not isomorphic. Responses in early visual areas are likely due to differential feedback processes from the dorsal and ventral streams. We, thus, conclude that surface color representation depends on a context-sensitive network of multiple distributed processes in the cortex.

Funding: This work was supported by a grant from the Agence Nationale de la Recherche to FD (ANR-11-JSH-20021) and by LABEX CORTEX (ANR-11-LABX-0042) to KK and HK and ANR-11-BSV4-501, CORE-NETS (H.K.), ANR-14-CE13-0033, ARCHI-CORE (H.K.), ANR-15-CE32-0016, CORNET (H.K.). The Grenoble MRI facility IRMaGe was partly funded by the French program

‘Investissement d’Avenir’ run by the Agence Nationale pour la Recherche (ANR-11-INBS-0006).

Acknowledgements: The authors thank Elisabeth Baumgartner, Karl Gegenfurtner, Stewart Shipp and Andrew Coia for comments and criticisms on a previous version of this manuscript. They also thank Mathieu Ruiz for help with data analysis and the technical staff at the Grenoble MRI facility IRMaGe.

References

Angelucci, A., and P. C. Bressloff. 2006. "Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons." *Prog Brain Res* 154:93-120. doi: 10.1016/S0079-6123(06)54005-1.

Anstis, S. 2010. "Visual filling-in." *Curr Biol* 20 (16):R664-6. doi: 10.1016/j.cub.2010.06.029.

Bartels, A., and S. Zeki. 2000. "The architecture of the colour centre in the human visual brain: new results and a review." *Eur J Neurosci* 12 (1):172-93.

Bellot, E., V. Coizet, J. Warnking, K. Knoblauch, E. Moro, and M. Dojat. 2016. "Effects of aging on low luminance contrast processing in humans." *Neuroimage* 139:415-426. doi: 10.1016/j.neuroimage.2016.06.051.

Bouvier, S. E., K. S. Cardinal, and S. A. Engel. 2008. "Activity in visual area V4 correlates with surface perception." *J Vis* 8 (7):28 1-9. doi: 10.1167/8.7.28.

Brainard, D. H. 1997. "The Psychophysics Toolbox." *Spat Vis* 10 (4):433-6.

Brouwer, G. J., and D. J. Heeger. 2009. "Decoding and reconstructing color from responses in human visual cortex." *J Neurosci* 29 (44):13992-4003. doi: 10.1523/JNEUROSCI.3577-09.2009.

Brown, H., and K. J. Friston. 2012. "Free-energy and illusions: the Cornsweet effect." *Front Psychol* 3:43. doi: 10.3389/fpsyg.2012.00043.

Caplovitz, G. P., and P. U. Tse. 2007. "V3A processes contour curvature as a trackable feature for the perception of rotational motion." *Cereb Cortex* 17 (5):1179-89. doi: 10.1093/cercor/bhl029.

Castaldi, E., F. Frijia, D. Montanaro, M. Tosetti, and M. C. Morrone. 2013. "BOLD human responses to chromatic spatial features." *Eur J Neurosci* 38 (2):2290-9. doi: 10.1111/ejn.12223.

Chevreul, M.E. . 1839. *De la loi du contraste simultané des couleurs et de l'assortiment des objets coloré.* Paris: Hachette Livre BNF (éd. 2012).

Coia, A. J., and M. A. Crognale. 2017. "Contour adaptation reduces the spreading of edge induced colors." *Vision Res*. doi: 10.1016/j.visres.2017.01.009.

Coia, A. J., C. Jones, C. S. Duncan, and M. A. Crognale. 2014. "Physiological correlates of watercolor effect." *J Opt Soc Am A Opt Image Sci Vis* 31 (4):A15-22. doi: 10.1364/JOSAA.31.000A15.

Cornelissen, F. W., A. R. Wade, T. Vladusich, R. F. Dougherty, and B. A. Wandell. 2006. "No functional magnetic resonance imaging evidence for brightness and color filling-in in early human visual cortex." *J Neurosci* 26 (14):3634-41. doi: 10.1523/JNEUROSCI.4382-05.2006.

Daniel, P. M., and D. Whitteridge. 1961. "The representation of the visual field on the cerebral cortex in monkeys." *J Physiol* 159:203-21.

De Weerd, P., R. Gattass, R. Desimone, and L. G. Ungerleider. 1995. "Responses of cells in monkey visual cortex during perceptual filling-in of an artificial scotoma." *Nature* 377 (6551):731-4. doi: 10.1038/377731a0.

Derrington, A. M., J. Krauskopf, and P. Lennie. 1984. "Chromatic mechanisms in lateral geniculate nucleus of macaque." *J Physiol* 357:241-65.

Devinck, F., P. Gerardin, M. Dojat, and K. Knoblauch. 2014. "Spatial selectivity of the watercolor effect." *J Opt Soc Am A Opt Image Sci Vis* 31 (4):A1-6. doi: 10.1364/JOSAA.31.0000A1.

Devinck, F., and K. Knoblauch. 2012. "A common signal detection model accounts for both perception and discrimination of the watercolor effect." *J Vis* 12 (3). doi: 10.1167/12.3.19.

DeYoe, E. A., G. J. Carman, P. Bandettini, S. Glickman, J. Wieser, R. Cox, D. Miller, and J. Neitz. 1996. "Mapping striate and extrastriate visual areas in human cerebral cortex." *Proc Natl Acad Sci U S A* 93 (6):2382-6.

Dupont, P., B. De Bruyn, R. Vandenberghe, A. M. Rosier, J. Michiels, G. Marchal, L. Mortelmans, and G. A. Orban. 1997. "The kinetic occipital region in human visual cortex." *Cereb Cortex* 7 (3):283-92.

Engel, S. A., D. E. Rumelhart, B. A. Wandell, A. T. Lee, G. H. Glover, E. J. Chichilnisky, and M. N. Shadlen. 1994. "fMRI of human visual cortex." *Nature* 369 (6481):525. doi: 10.1038/369525a0.

Friedman, H. S., H. Zhou, and R. von der Heydt. 2003. "The coding of uniform colour figures in monkey visual cortex." *J Physiol* 548 (Pt 2):593-613. doi: 10.1113/jphysiol.2002.033555.

Friston, K. J. 2011. "Functional and effective connectivity: a review." *Brain Connect* 1 (1):13-36. doi: 10.1089/brain.2011.0008.

Friston, K. J., L. Harrison, and W. Penny. 2003. "Dynamic causal modelling." *Neuroimage* 19 (4):1273-302.

Friston, K., and S. Kiebel. 2009. "Predictive coding under the free-energy principle." *Philos Trans R Soc Lond B Biol Sci* 364 (1521):1211-21. doi: 10.1098/rstb.2008.0300.

Gerardin, P., F. Devinck, M. Dojat, and K. Knoblauch. 2014. "Contributions of contour frequency, amplitude, and luminance to the watercolor effect estimated by conjoint measurement." *J Vis* 14 (4). doi: 10.1167/14.4.9.

Gerardin, P., Z. Kourtzi, and P. Mamassian. 2010. "Prior knowledge of illumination for 3D perception in the human brain." *Proc Natl Acad Sci U S A* 107 (37):16309-14. doi: 10.1073/pnas.1006285107.

Hadjikhani, N., A. K. Liu, A. M. Dale, P. Cavanagh, and R. B. Tootell. 1998. "Retinotopy and color sensitivity in human visual cortical area V8." *Nat Neurosci* 1 (3):235-41. doi: 10.1038/681.

Haynes, J. D., and G. Rees. 2005. "Predicting the orientation of invisible stimuli from activity in human primary visual cortex." *Nat Neurosci* 8 (5):686-91. doi: 10.1038/nn1445.

Hazenberg, S. J., and R. van Lier. 2013. "Afterimage watercolors: an exploration of contour-based afterimage filling-in." *Front Psychol* 4:707. doi: 10.3389/fpsyg.2013.00707.

Holm, S. 1979. "A simple sequentially rejective multiple test procedure." *Scandinavian Journal of Statistics* 6:65-70.

Hong, S. W., and F. Tong. 2017. "Neural representation of form-contingent color filling-in in the early visual cortex." *J Vis* 17 (13):10. doi: 10.1167/17.13.10.

Jamalabadi, H., S. Alizadeh, M. Schonauer, C. Leibold, and S. Gais. 2016. "Classification based hypothesis testing in neuroscience: Below-chance level classification rates and overlooked statistical properties of linear parametric classifiers." *Hum Brain Mapp* 37 (5):1842-55. doi: 10.1002/hbm.23140.

Joachims, T. 1999. *Making large-scale SVM learning practical. Advances in Kernel Methods – Support Vector Learning.* MIT Press.: B. Schölkopf, C. Burges and A. Smola.

Johnson, E. N., M. J. Hawken, and R. Shapley. 2001. "The spatial transformation of color in the primary visual cortex of the macaque monkey." *Nat Neurosci* 4 (4):409-16. doi: 10.1038/86061.

Kamitani, Y., and F. Tong. 2005. "Decoding the visual and subjective contents of the human brain." *Nat Neurosci* 8 (5):679-85. doi: 10.1038/nn1444.

Knoblauch, K., and L.T. Maloney. 2008. "MLDS: Maximum likelihood difference scaling in R." *Journal of Statistical Software* 25 (2):1-26.

Knoblauch, K., and L.T. Maloney. 2012. "Modeling Psychophysical Data in R." *New York: Springer*.

Kok, P., and F. P. de Lange. 2014. "Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex." *Curr Biol* 24 (13):1531-5. doi: 10.1016/j.cub.2014.05.042.

Komatsu, H. 2006. "The neural mechanisms of perceptual filling-in." *Nat Rev Neurosci* 7 (3):220-31. doi: 10.1038/nrn1869.

Konen, C. S., and S. Kastner. 2008. "Two hierarchically organized neural systems for object information in human visual cortex." *Nat Neurosci* 11 (2):224-31. doi: 10.1038/nn2036.

Kourtzi, Z., and N. Kanwisher. 2001. "Representation of perceived object shape by the human lateral occipital complex." *Science* 293 (5534):1506-9. doi: 10.1126/science.1061133.

Kravitz, D. J., K. S. Saleem, C. I. Baker, L. G. Ungerleider, and M. Mishkin. 2013. "The ventral visual pathway: an expanded neural framework for the processing of object quality." *Trends Cogn Sci* 17 (1):26-49. doi: 10.1016/j.tics.2012.10.011.

Lee, T. S., and M. Nguyen. 2001. "Dynamics of subjective contour formation in the early visual cortex." *Proc Natl Acad Sci U S A* 98 (4):1907-11. doi: 10.1073/pnas.031579998.

Leopold, D. A., and N. K. Logothetis. 1996. "Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry." *Nature* 379 (6565):549-53. doi: 10.1038/379549a0.

Maloney, L. T., and J. N. Yang. 2003. "Maximum likelihood difference scaling." *J Vis* 3 (8):573-85. doi: 10:1167/3.8.5.

Markov, N. T., M. Ercsey-Ravasz, D. C. Van Essen, K. Knoblauch, Z. Toroczkai, and H. Kennedy. 2013. "Cortical high-density counterstream architectures." *Science* 342 (6158):1238406. doi: 10.1126/science.1238406.

Markov, N. T., P. Misery, A. Falchier, C. Lamy, J. Vezoli, R. Quilodran, M. A. Gariel, P. Giroud, M. Ercsey-Ravasz, L. J. Pilaz, C. Huissoud, P. Barone, C. Dehay, Z. Toroczkai, D. C. Van Essen, H. Kennedy, and K. Knoblauch. 2011. "Weight consistency specifies regularities of macaque cortical networks." *Cerebral Cortex* 21 (6):1254-72. doi: 10.1093/cercor/bhq201.

Mendola, J. D., I. P. Conner, S. Sharma, A. Bahekar, and S. Lemieux. 2006. "fMRI Measures of perceptual filling-in in the human visual cortex." *J Cogn Neurosci* 18 (3):363-75. doi: 10.1162/089892906775990624.

Michalareas, G., J. Vezoli, S. van Pelt, J-M. Schoffelen, H. Kennedy, and P. Fries. 2016. "Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual areas." *Neuron* 89 (2):384-397.

Nieuwenhuis, S., B. U. Forstmann, and E. J. Wagenmakers. 2011. "Erroneous analyses of interactions in neuroscience: a problem of significance." *Nat Neurosci* 14 (9):1105-7. doi: 10.1038/nn.2886.

Parkes, L. M., J. B. Marsman, D. C. Oxley, J. Y. Goulermas, and S. M. Wuerger. 2009. "Multivoxel fMRI analysis of color tuning in human primary visual cortex." *J Vis* 9 (1):1 1-13. doi: 10.1167/9.1.1.

Pelli, D. G. 1997. "The VideoToolbox software for visual psychophysics: transforming numbers into movies." *Spat Vis* 10 (4):437-42.

Peng, X., and D. C. Van Essen. 2005. "Peaked encoding of relative luminance in macaque areas V1 and V2." *J Neurophysiol* 93 (3):1620-32. doi: 10.1152/jn.00793.2004.

Pinheiro, J.C., and D.M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

Pinna, B., G. Brelstaff, and L. Spillmann. 2001. "Surface color from boundaries: a new 'watercolor' illusion." *Vision Res* 41 (20):2669-76.

Pinna, B., and S. Grossberg. 2005. "The watercolor illusion and neon color spreading: a unified analysis of new cases and neural mechanisms." *J Opt Soc Am A Opt Image Sci Vis* 22 (10):2207-21.

Press, W. A., A. A. Brewer, R. F. Dougherty, A. R. Wade, and B. A. Wandell. 2001. "Visual areas and spatial summation in human visual cortex." *Vision Res* 41 (10-11):1321-32.

Ratliff, F., and L. Sirovich. 1978. "Equivalence classes of visual stimuli." *Vision Res* 18 (7):845-51.

RCoreTeam. 2015. "R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria." *ISBN 3-900051-07-0*.

Rigoux, L., K. E. Stephan, K. J. Friston, and J. Daunizeau. 2014. "Bayesian model selection for group studies - revisited." *Neuroimage* 84:971-85. doi: 10.1016/j.neuroimage.2013.08.065.

Roe, A. W., and D. Y. Ts'o. 1999. "Specificity of color connectivity between primate V1 and V2." *J Neurophysiol* 82 (5):2719-30.

Sakai, K., E. Watanabe, Y. Onodera, I. Uchida, H. Kato, E. Yamamoto, H. Koizumi, and Y. Miyashita. 1995. "Functional mapping of the human colour centre with echo-planar magnetic resonance imaging." *Proc Biol Sci* 261 (1360):89-98. doi: 10.1098/rspb.1995.0121.

Sapountzis, P., D. Schluppeck, R. Bowtell, and J. W. Peirce. 2010. "A comparison of fMRI adaptation and multivariate pattern classification analysis in visual cortex." *Neuroimage* 49 (2):1632-40. doi: 10.1016/j.neuroimage.2009.09.066.

Sasaki, Y., and T. Watanabe. 2004. "The primary visual cortex fills in color." *Proc Natl Acad Sci U S A* 101 (52):18251-6. doi: 10.1073/pnas.0406293102.

Schluppeck, D., and S. A. Engel. 2002. "Color opponent neurons in V1: a review and model reconciling results from imaging and single-unit recording." *J Vis* 2 (6):480-92. doi: 10.1167/2.6.5.

Sereno, M. I., A. M. Dale, J. B. Reppas, K. K. Kwong, J. W. Belliveau, T. J. Brady, B. R. Rosen, and R. B. Tootell. 1995. "Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging." *Science* 268 (5212):889-93.

Stephan, K. E., W. D. Penny, J. Daunizeau, R. J. Moran, and K. J. Friston. 2009. "Bayesian model selection for group studies." *Neuroimage* 46 (4):1004-17. doi: 10.1016/j.neuroimage.2009.03.025.

Stephan, K. E., W. D. Penny, R. J. Moran, H. E. den Ouden, J. Daunizeau, and K. J. Friston. 2010. "Ten simple rules for dynamic causal modeling." *Neuroimage* 49 (4):3099-109. doi: 10.1016/j.neuroimage.2009.11.015.

Stettler, D. D., A. Das, J. Bennett, and C. D. Gilbert. 2002. "Lateral connectivity and contextual interactions in macaque primary visual cortex." *Neuron* 36 (4):739-50.

Takemura, H., A. Rokem, J. Winawer, J. D. Yeatman, B. A. Wandell, and F. Pestilli. 2015. "A Major Human White Matter Pathway Between Dorsal and Ventral Visual Cortex." *Cereb Cortex*. doi: 10.1093/cercor/bhv064.

Tootell, R. B., K. Nelissen, W. Vanduffel, and G. A. Orban. 2004. "Search for color 'center(s)' in macaque visual cortex." *Cereb Cortex* 14 (4):353-63.

Tootell, R. B., J. B. Reppas, K. K. Kwong, R. Malach, R. T. Born, T. J. Brady, B. R. Rosen, and J. W. Belliveau. 1995. "Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging." *J Neurosci* 15 (4):3215-30.

Van Essen, D. C., W. T. Newsome, and J. H. Maunsell. 1984. "The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability." *Vision Res* 24 (5):429-48.

van Lier, R., M. Vergeer, and S. Anstis. 2009. "Filling-in afterimage colors between the lines." *Curr Biol* 19 (8):R323-4. doi: 10.1016/j.cub.2009.03.010.

Varin, D. 1971. "Fenomeni di contrasto e diffusione cromatica nell'organizzazione spaziale del campo percettivo." *Rivista di Psicologia* 65:101-128.
Von der Heydt, R., and R. Pierson. 2006. "Dissociation of color and figure-ground effects in the watercolor illusion." *Spat Vis* 19 (2-4):323-40.
Warnking, J., M. Dojat, A. Guerin-Dugue, C. Delon-Martin, S. Olympieff, N. Richard, A. Chehikian, and C. Segebarth. 2002. "fMRI retinotopic mapping--step by step." *Neuroimage* 17 (4):1665-83.
Yang, J. N., N. M. Szeverenyi, and D. Ts'o. 2008. "Neural resources associated with perceptual judgment across sensory modalities." *Cereb Cortex* 18 (1):38-45. doi: 10.1093/cercor/bhm029.
Zipser, K., V. A. Lamme, and P. H. Schiller. 1996. "Contextual modulation in primary visual cortex." *J Neurosci* 16 (22):7376-89.
Zweig, S., G. Zurawel, R. Shapley, and H. Slovin. 2015. "Representation of Color Surfaces in V1: Edge Enhancement and Unfilled Holes." *J Neurosci* 35 (35):12103-15.