

Invariance of surface color representations across illuminant changes in the human cortex



Michael M. Bannert^{a,b,c,d,e,*}, Andreas Bartels^{a,b,c,d,*}

^a Vision and Cognition Lab, Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, 72076 Tübingen, Germany

^b Bernstein Center for Computational Neuroscience, 72076 Tübingen, Germany

^c Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany

^d Department of Psychology, University of Tübingen, 72076 Tübingen, Germany

^e International Max Planck Research School for Cognitive and Systems Neuroscience, 72076 Tübingen, Germany

ARTICLE INFO

Keywords:

Color constancy

fMRI

V1

V4

Surface perception

ABSTRACT

A central problem in color vision is that the light reaching the eye from a given surface can vary dramatically depending on the illumination. Despite this, our color percept, the brain's estimate of surface reflectance, remains remarkably stable. This phenomenon is called color constancy. Here we investigated which human brain regions represent surface color in a way that is invariant with respect to illuminant changes. We used physically realistic rendering methods to display natural yet abstract 3D scenes that were displayed under three distinct illuminants. The scenes embedded, in different conditions, surfaces that differed in their surface color (i.e. in their reflectance property). We used multivariate fMRI pattern analysis to probe neural coding of surface reflectance and illuminant, respectively. While all visual regions encoded surface color when viewed under the same illuminant, we found that only in V1 and V4 surface color representations were invariant to illumination changes. Along the visual hierarchy there was a gradient from V1 to V4 to increasingly encode surface color rather than illumination. Finally, effects of a stimulus manipulation on individual *behavioral* color constancy indices correlated with neural encoding of the *illuminant* in hV4. This provides neural evidence for the Equivalent Illuminant Model. Our results provide a principled characterization of color constancy mechanisms across the visual hierarchy, and demonstrate complementary contributions in early and late processing stages.

1. Introduction

Color constitutes a fundamental quality of visual experience, and supports a large variety of behavioral tasks (Mollon, 1989). However, the fact that the light reflected from surfaces depends both on the surface color (i.e. its reflectance) and the color of the incident light (Land and McCann, 1971) poses a challenging problem to the visual system. It is therefore impossible to know the reflectance of a surface without any knowledge of the illumination. As numerous psychophysical studies have documented, however, the perception of surface color is fairly robust even in the face of changes in illumination. This property of the visual system is referred to as “color constancy”. It is unclear how the human brain transforms the highly ambiguous incoming color signals to create surface color representations that are stable across illuminants. What factors at the neural level are involved in color constancy?

Early investigations demonstrated the involvement of area V4 in color perception in monkeys (Wild et al., 1985; Zeki, 1983) and inspired neuroimaging studies suggesting a similar role for human V4 (Bartels and Zeki, 2000; Beauchamp et al., 1999; Lueck et al., 1989). Also the functional organization of color responses in this area was shown to reflect perceptually relevant stimulus dimensions in both non-human (Conway and Tsao, 2009; Kusunoki et al., 2006; Li et al., 2014) and human primates (Brouwer and Heeger, 2013, 2009).

Human lesion studies have accordingly implicated a connection between area V4 and achromatopsia (but also form vision deficits) (Bouvier and Engel, 2005). As for selective color constancy impairments, evidence appears less conclusive with some work suggesting a link with V4 (Clarke et al., 1998; Kennard et al., 1995) while different research highlights the involvement of other areas (Rüttiger et al., 1999), including V1 (Kenridge et al., 2007).

* Corresponding authors. Vision and Cognition Lab, Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Str. 25, 72076 Tübingen, Germany.
E-mail addresses: michael.bannert@tuebingen.mpg.de (M.M. Bannert), andreas.bartels@tuebingen.mpg.de (A. Bartels).

Human neuroimaging indeed shows increased responses to color already in early areas V1 and V2 (Bartels and Zeki, 2000; Beauchamp et al., 1999; Engel et al., 1997). But also in non-human primates, the chromatic context modulation of neural color tuning (Wachtler et al., 2003) and the double opponency of neurons (Conway, 2001; Conway and Livingstone, 2006; Johnson et al., 2008, 2001) outline possible early color constancy mechanisms in V1.

While prior monkey studies shed some light on neural responses encoding perceptual (i.e. color constant) versus physical color properties in isolated regions, to our knowledge no prior study manipulated reflectance and illumination spectra to examine regional encoding of perceptually constant colors or of illuminant systematically across the whole ventral visual pathway, neither in monkey nor human brains.

We used multi-voxel pattern analysis as a test for color constancy: if a neural surface color representation is invariant across illumination changes, distinctions between representations of different surface colors should generalize across these changes. Using physically realistic rendering methods we showed subjects 3D scenes rendered in three conditions that simulated three different illuminants. Surfaces that differed in their surface color were embedded in these scenes. We designed our experiment in this way to achieve higher ecological validity: color vision in real life occurs in 3D environments, and it is well established that some surface color cues depend on 3D scene structure (Maloney, 1999; Radonjic et al., 2015). Participants performed an attention task (that was independent of illuminant or color) on these surfaces during fMRI recording.

To test our hypotheses, we trained classifiers to discriminate BOLD responses to two surfaces (“blue” and “yellow”) under two out of three illuminations (e.g. “neutral”, and “blue”) and tested them on new BOLD responses measured in the third illumination condition (e.g. “yellow”), which was not part of the training set. This analysis showed that activity patterns in V1 and V4 α encoded surface colors in a way that generalized across illumination conditions, i.e. in a color-constant way.

Furthermore, we tested a prediction from equivalent illuminant models of color constancy. We hypothesized that the neural accuracy of encoding the *illuminant* of a scene predicts the behavioral accuracy of constant color perception. We collected behavioral color constancy indices, including a cue conflict manipulation that abolishes behavioral color constancy. We collected fMRI data for the same stimuli. In accord with the equivalent illuminant model, we found that the behavioral effect of the cue conflict manipulation on color constancy could be predicted from neural decoding of the illuminant in hV4.

Lastly, we examined how visual areas interpret two different surfaces that reflect the same light because they are presented under different illuminations. These surfaces were perceived as having distinct colors, but emitted the same light. These stimuli can be discriminated on the basis of their surface reflectance or illumination. Our analysis revealed that higher visual regions hV4, VO1, V4 α weighted the difference in surface reflectance more strongly than earlier visual areas.

In sum, the results provide a detailed account of the contributions of different visual areas to color constancy.

2. Materials and methods

2.1. Participants

Our sample consisted of 20 healthy observers (15 female, 5 male) from the Tübingen University community between the ages of 19–35 (mean age: 24.5). All participants had normal color vision as ascertained with Ishihara plates (Ishihara, 2011). They provided written informed consent to participation in the experiment prior to the first session. The local ethics committee of the University Hospital Tübingen approved the study. Data from the fMRI main experiment of subject 12 could not be analyzed due to a data handling error. We therefore discarded this subject's dataset altogether, allowing us to use data of 19 subjects for the reported analyses.

2.2. Stimuli

2.2.1. Rendering method

We used the RenderToolbox3 (Heasley et al., 2014) to devise stimuli very similar to those used previously in a typical psychophysics experiment about color constancy (Xiao et al., 2012). The RenderToolbox3 provides a MATLAB-based framework for the development of stimuli with the 3D modeling software Blender 2.73 (www.blender.org) and the rendering software Mitsuba (www.mitsuba-renderer.org). The toolbox enables the user to control material properties (like reflectance) and the spectral power distributions (SPDs) of light sources within a 3D scene and creates 2D images of that scene based on a physically accurate rendering algorithm. The domains of reflectance functions and SPDs ranged from 380 nm to 730 nm and were discretized in steps of 10 nm.

2.2.2. Scenes, illuminants, and surfaces of interest

We rendered complex yet abstract three dimensional scenes, similar to those used in prior behavioral studies on color constancy (Xiao et al., 2012). The key motivation for using complex scenes was evidence that color constancy benefits from complex surroundings, presumably as the latter provides better cues to estimate the illuminant (Maloney, 1999; Radonjic et al., 2015). We embedded several surface patches in these scenes (see Fig. 1a). The patches had, in different experimental conditions, two different surface reflectance functions. These functions were chosen such that the surfaces appeared blue and yellow, respectively, under “neutral”, daylight illumination, which was approximated by the standard illuminant D65 (Fig. 1a). From here onward we refer to these surfaces of interest simply as “surfaces” or “surface patches”. The full scene (including the surfaces), was rendered under a total of three illuminants: the standard daylight illuminant D65 (Fig. 1a), plus two more illuminants located on the black body curve at correlated color temperatures of 10,925 K (Fig. 1b) and 4500 K (Fig. 1c). For simplicity, the first illuminant is referred to as the “neutral” illuminant, the second and third as “blue” and “yellow” illuminants due to their appearances compared to daylight. The reflectance functions of the two surfaces were chosen such that the blue surface under the yellow illuminant reflected the same light as the yellow surface under the blue illuminant (compare SPDs of reflected lights in Fig. 1b on the right and Fig. 1c on the left). Together, the two surfaces and three illumination conditions constitute six conditions.

The colored surfaces appeared predominantly in the left visual hemifield for odd-numbered participants (Fig. 1) and in the right hemifield in even-numbered participants (not shown). Three surface patches were shown in one hemifield and one in the other. We chose this stimulus design to exploit the retinotopic organization in visual cortex, which allowed us to test for hemispheric lateralization of BOLD responses.

Note that each rendered scene (including surface patches) provides cues on its illuminant: the illuminant illuminates the whole scene. However, and importantly, the scene excluding the surface patches does not provide any information on the surface color. Surface color cues are hence exclusively provided by the surfaces (within a given illuminant condition), and by the *combination* of the scene and the surfaces (across illuminants): under a given illuminant, the light emitted from the surfaces alone does differentiate between the two surface colors. But this is not the case across illuminants, as under the blue illuminant the yellow surface emits identical light as the blue surface under yellow illumination. To achieve color constancy, the brain needs to integrate information from the scene *and* the surface to infer the reflectance property of the surface, as it requires information of the illuminant and of the light reflected from the surface to infer its color.

Fig. 2 shows the chromaticity coordinates of the two surfaces under each of the three illuminants. Average patch luminance in the neutral condition was 284.4 cd/m². Blue and yellow surfaces were matched in luminance. We also matched stimulus luminance across illumination conditions. The Michelson contrast between patches and the immediately surrounding (i.e., within 5 pixels) background surface was 0.179.

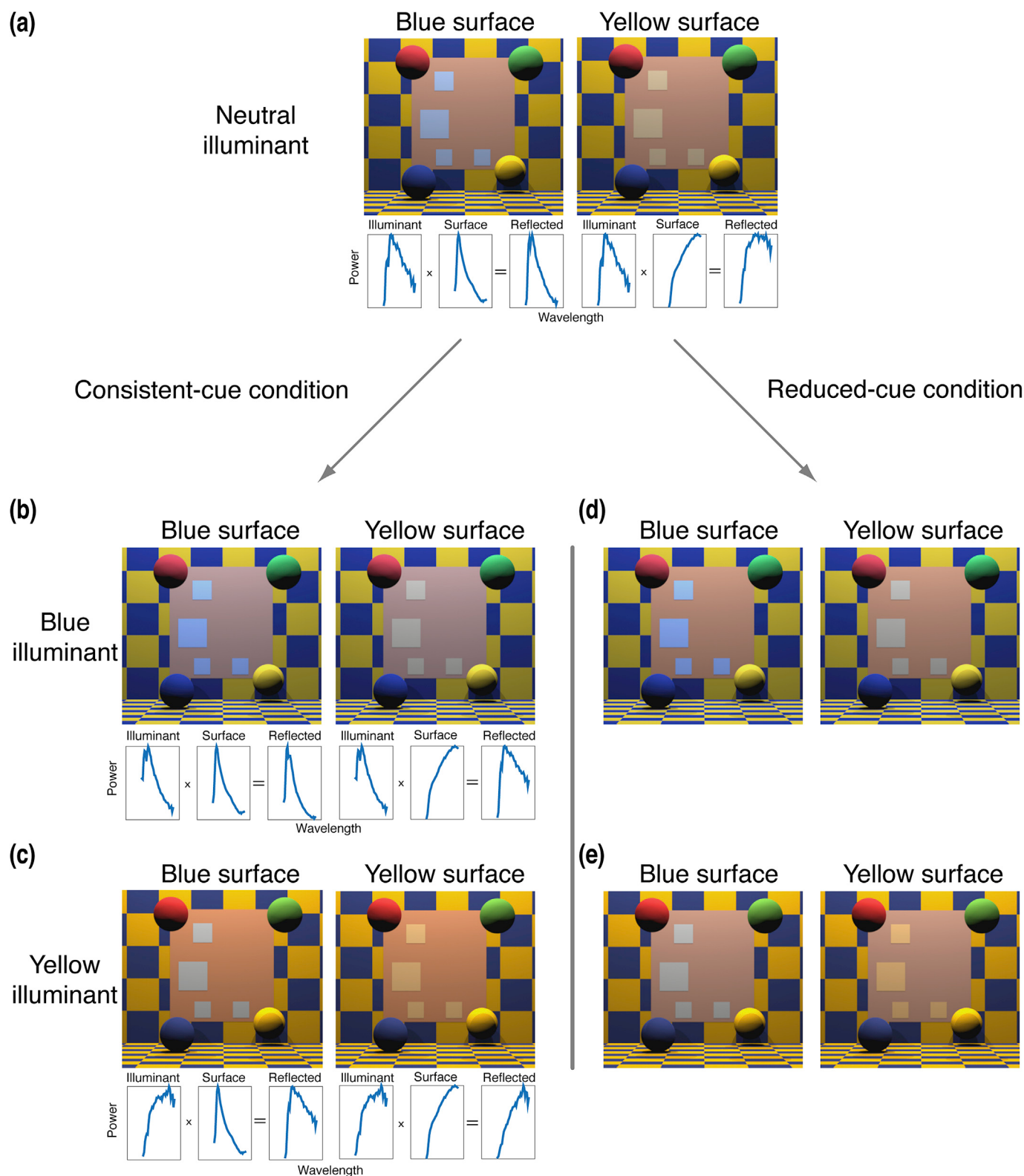


Fig. 1. Experimental Procedure and Stimuli. **(a)** Rendered 3D scenes contained four surfaces that were either blue or yellow. In even-numbered subjects the location of the four surfaces was mirrored horizontally at the center (not shown). Note that 3 surfaces fell in one hemifield, 1 in the other. Plots below each scene image show SPDs for the illuminant, the reflectance functions for the surface, and light reflected from the surfaces. **(b, c)** Same scenes as in **(a)** but rendered under blue and yellow illumination, respectively. Note that the SPDs of the lights reflected from the yellow surface under blue illumination and from the blue surface under yellow illumination are the same. **(e, f)** Same as in **(b)** and **(c)** except this time the background rectangle was replaced with a different surface that reflected the same light as the original one under neutral illumination shown in **(a)**. Note that this figure serves as illustration only, and does not allow judgment of actual color constancy effects produced by the stimuli shown in experimental conditions. Robustness of color constancy in experimental conditions is quantified in Fig. 5.

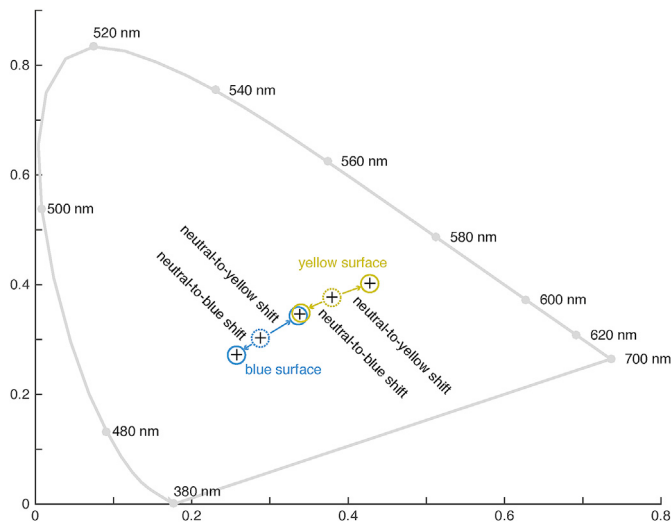


Fig. 2. Stimulus chromaticities. Stimulus coordinates (x, y) in the CIE 1931 chromaticity plane. Blue (yellow) circles mark chromaticities of the blue (yellow) surface. Dotted circles indicate chromaticities under D65 illumination. Arrows represent chromaticity shifts induced by changes in illumination. Note that a chromaticity shift from neutral to blue for the yellow surface results in the same chromaticity as a shift from neutral to yellow for the blue surface as indicated by overlapping blue and yellow circles.

2.2.3. Cue conflict conditions

In addition to the above, we introduced cue conflict conditions shown to strongly impair color constancy judgments (Delahunt and Brainard, 2004; Xiao et al., 2012). The manipulation consists in replacing the central large background rectangle on which the blue or yellow colored surfaces appeared with a background that reflected the same light under each illuminant as the original background under neutral illumination (Fig. 1d and e). We refer to the condition involving the original background as the *consistent-cue* condition and to the condition with the replaced background as the *reduced-cue* condition (following Xiao et al., 2012). The manipulation was applied for the blue and yellow illuminant and both surfaces, yielding four cue conflict conditions – the neutral cue-conflict condition was identical to the neutral consistent-cue condition. In total, we had ten different conditions.

To make sure that classification of surface colors was driven only by differences between the chromaticities of the two surfaces and not by luminance differences, we took two measures: first, we equated luminances across both surface colors and all illumination conditions. This was achieved by setting the images to the average luminance values across the ten conditions on a pixel-by-pixel basis. Second, one half of the stimuli was presented with luminance increased and the other half with luminance reduced by 10%. This ensured that classifiers generalized across possible differences in luminance and made discriminations based on chromatic differences. To accomplish this, we converted the RGB values of all images to XYZ tri-stimulus space using the transformation matrix obtained from the display calibration. We then calculated the mean luminance (Y) across all images. We converted each XYZ image to CIE xyY space to separate chromaticity from luminance components and applied the mean luminance vector to every image (including the two 10% luminance modulations). The images were subsequently transformed back to RGB space for presentation on the gamma-corrected displays.

2.3. fMRI experiment

Participants lay supine in the scanner and viewed the scene images via a mirror mounted on top of the head coil. Stimuli were presented against a screen located at the end of the scanner bore using a gamma-calibrated projector (NEC PE401H, CalibrateMonSpd.m function from Psychtoolbox, spectroradiometer by Photo Research PR-670). The size of

each image on the projection screen was 16.8° and 15° of visual angle in horizontal and vertical directions, respectively, at a resolution of 1024×768 pixels.

2.3.1. Stimulus presentation

The sequence of trials is shown in Fig. 3. Each scene appeared four times for 1.5 s within a stimulus block in alternation with a luminance matched color mask that lasted for 1 s, leading to a block-duration of 10 s. Color masks consisted of three RGB layers created independently from 1/f noise. All stimuli were presented using MATLAB and Psychtoolbox (Kleiner et al., 2007). Stimulus blocks appeared on the screen in a pseudo-randomized sequence that made sure that every pair-wise transition between conditions was equally likely across the whole experiment (Brooks, 2012). All ten conditions were presented 40 times each in a total of eight runs. Each run lasted 8 min 20 s plus 8.7 s for 11 dummy scans and 8 s extra scan time.

2.3.2. Task

Participants maintained fixation on the fixation cross in the center of the screen while paying attention to the colored surfaces surrounding it. Their task was to respond via button press whenever they detected a target event. Targets were a temporary decrease in the luminance of the white fixation cross or a temporary change in the number of colored surfaces. One of the surfaces at the 11, 9, 7, or 5 o'clock positions sometimes disappeared for a short period or an additional surface appeared at the 1 or 3 o'clock position. The target event lasted 0.5 s in each case. Each of the six surface events occurred in 3.75% of the trials while fixation cross events occurred in 22.5% of the trials yielding a total target probability of 45% ($= 22.5\% + 6 \text{ locations} \times 3.75\%$) for each condition. To increase motivation, participants received feedback about their performance at the end of each run.

2.4. Retinotopic mapping & ROI definition

Visual areas V1–V3, hV4, and VO1 were localized using standard retinotopic mapping techniques (Sereno et al., 1995). In brief, observers fixated in the middle of the screen while attending a flickering black and white checkerboard visible through a wedge-shaped aperture on a gray background. Check sizes increased logarithmically from the center to the visual periphery to account for cortical magnification. The wedge subtended the entire screen within an angle of 90° and rotated with a period of 55.64 s in clockwise or counter-clockwise direction. Participants viewed ten cycles of this stimulus in four polar mapping runs with stimulus direction alternating between runs. The functional data were analyzed using *Freesurfer* software (<http://surfer.nmr.mgh.harvard.edu/>) to obtain polar maps on a flattened 2D reconstruction of individual brains that allowed for the detailed demarcation of single visual areas.

In addition to retinotopic ROIs we also created a group ROI located anterior to the retinotopically defined ROIs based on the results of our within-illuminant searchlight analysis (see section “fMRI pattern classification: searchlight analysis” in Results). This ROI was defined as the smallest sphere that encompassed an entire cluster of information located in a brain region that has previously been labeled V4 α (Barbur and Spang, 2008; Bartels and Zeki, 2000). We hence refer to this region as putative V4 α (or pV4 α for short). The MNI coordinates of the center of the sphere were $x = 39$, $y = -58$, $z = -11$ and its radius was 6 mm. Importantly, the searchlight analysis used to define pV4 α was independent from the subsequent classification tests carried out on this ROI.

2.5. fMRI scan details

We used a 64-channel coil to record functional activity at 3 T magnetic field strength (Siemens Prisma) with a GRAPPA accelerated T2*-weighted 4-fold parallel imaging sequence (GRAPPA factor 2, 0.87 s repetition time (TR), 57° flip angle, 30 ms echo time (TE), 56 slices without gaps, 96×96 acquisition matrix, 2 mm isotropic voxel size). A

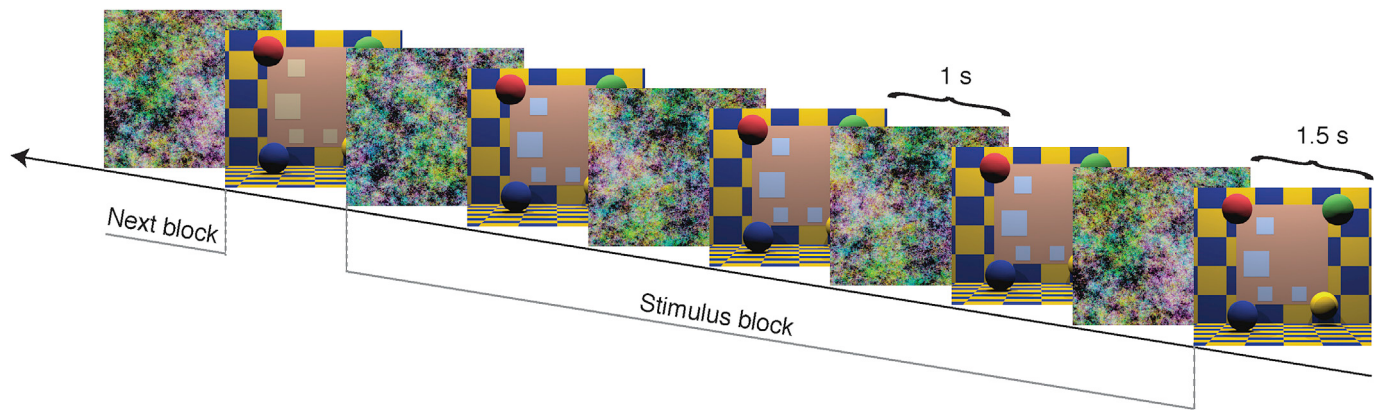


Fig. 3. Stimulus Presentation Sequence. Each scene was presented four times for 1.5 s, each time followed by a random color mask shown for 1 s. Observers were instructed to press a button whenever the fixation cross was dimmed or when the number of colored surfaces changed.

T1-weighted MP-RAGE sequence was used to acquire an anatomical image of each subject's brain at a resolution of $1 \times 1 \times 1 \text{ mm}^3$. To correct for magnetic field inhomogeneity, we measured a Gradient Echo field map in each fMRI session.

2.6. fMRI data analysis

2.6.1. fMRI data preprocessing

Preprocessing was carried out using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>). We applied motion correction and unwarping (using the voxel displacement maps obtained from the Gradient Echo field map) to our functional data, corrected them for slice acquisition time, and co-registered them to the anatomical scan. The anatomical scans were segmented and normalized to MNI space along with the functional data. The co-registered functional data from the retinotopic mapping experiment were further smoothed with a 4 mm (FWHM) Gaussian kernel.

2.6.2. fMRI pattern estimation

To estimate the patterns of fMRI responses to our stimuli under various conditions, we modeled the normalized, unsmoothed voxel time series with GLMs. Every stimulus block (Fig. 3) was modeled separately as a boxcar function shifted forward in time by 5 s to account for the hemodynamic lag. Additional regressors modeled the estimated motion parameters. For each time step voxel values were scaled by the global mean value. Instead of applying SPM's frequency cutoff, the resulting series of beta estimates were temporally filtered by removing linear and quadratic trends (and intercept) from each voxel per run. Finally, we z-scored the beta estimates for each voxel and replaced values above 2 with 2s and values below -2 with -2 s to handle outliers as recommended by the LIBLINEAR authors (see below).

2.6.3. fMRI pattern analysis

All classification analyses were carried out using leave-one-run-out cross-validation, and all inference was based on bootstrapping and family-wise error correction (FWE) (for inference see section on statistical inference below). In detail, classification accuracy was calculated by taking the sum of correct percentages across validation folds weighted by the number of predictions made in that run. If training sets were unbalanced (i.e., contained more samples from one category than the other), we imputed the missing values by setting them to the average pattern of that category. We used the LIBLINEAR implementation (<https://www.csie.ntu.edu.tw/~cjlin/liblinear/>) of SVM (Fan et al., 2008). The training algorithm determined the optimal value for C by performing 3-fold cross-validation on training data and selecting the value that minimized prediction error. C is a penalty parameter that controls the trade-off between training set error and generalization to prevent overfitting.

Candidate values for C were powers of 2: 2^{-18} , 2^{-17} , ..., 2^1 .

2.6.3.1. Within illuminant classification. To provide a benchmark of decoding accuracies for subsequent analyses, we first tested if the two surface colors could be predicted from fMRI activity patterns when training and test set data came from the same illuminant conditions. In this case above-chance classification does not specifically depend on perceived surface color but can be achieved simply on the basis of different wavelength compositions of the lights the surfaces reflect. Discriminability of surface color within the same illumination conditions is rather a prerequisite if classification is expected to work when training and testing are performed on different illumination conditions. This analysis can be regarded as a replication of previous research (e.g., Brouwer and Heeger, 2009) but extends it to the more complex context of our 3D rendered scene that embeds the surfaces of interest. We proceeded by training classifiers to distinguish between blue and yellow surfaces under each two pairs of illuminants (neutral/blue-consistent-cue, neutral/yellow-consistent-cue, blue-consistent-cue/yellow-consistent-cue) and tested them on responses to stimuli from the same two conditions (recorded in the withheld run) (Fig. 4a). The average across all three possible combinations was taken as the dependent variable in this analysis.

We provided the algorithm with vectors of voxel response patterns (one for each trial), and for each one label corresponding to the surface color (regardless of illuminant) that was presented to the observer when that measurement was made. The algorithm was “blind” with regards to the illumination that the measured pattern came from. The classifier thus had to treat any variation due to our luminance modulation or illumination as within-class variation. This approach helps the classifier select especially those voxels that are more likely to represent color in a constant way while avoiding voxels that are driven by luminance differences. At the same time, this approach allowed us to exploit as much of our data as possible in our analyses. This training procedure was chosen for all three analyses (Fig. 4a–c) to allow for comparison.

The critical difference between the classifications within-illuminants and across-illuminants is this: In within-illuminant classification the test data from the withheld run came from the same two illuminant conditions while in across-illuminant classification the test data came from the third illuminant condition.

The reason why we used data of two illuminants also for training in our control (“within illuminant”) analysis is described in the following. For “within illuminant” we trained a classifier to distinguish “blue” and “yellow” surfaces that were viewed under two illuminants (e.g. “neutral” and “blue”). Testing was done for the same two illuminants on trials of a left-out run. Hence, for ROIs that did not have illuminant-invariant representations, the classifier would learn two patterns for “blue”

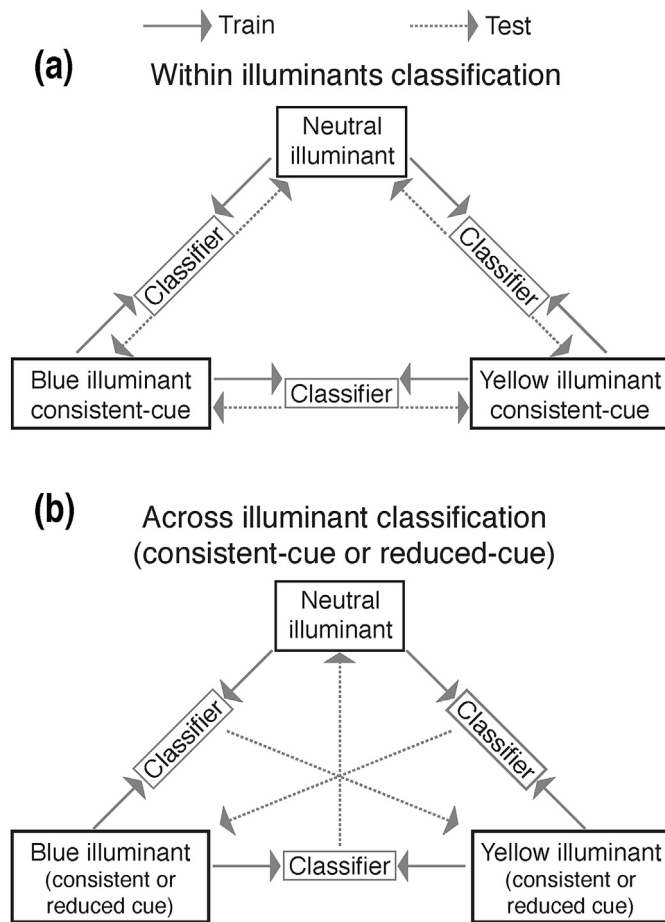


Fig. 4. Classification Analyses. Illustration of the first three classification analyses performed on fMRI response patterns. These classifications aimed to distinguish between blue and yellow surface conditions, within or across varying illumination conditions, respectively. **(a)** Three classifiers were trained to distinguish responses to blue and yellow surfaces under two different illuminants and tested on responses to stimuli from the same illumination conditions that were not used for training. No generalization across illuminations was required. Results were averaged across the three analyses. **(b)** Training as in (a), but this time classifiers were tested on responses to stimuli from the third illumination condition that was not included for training. The analysis for (b) was carried out separately for consistent-cue and for reduced-cue stimuli, respectively.

surface (one for each illuminant) and another two for “yellow” surface, and perform well on testing using trials taken from either illuminant. Testing on a new (untrained) illuminant would fail. The advantage of this training procedure on two illuminants, and the reason why we used it, was that for ROIs that did have illuminant-invariant representations, training would encourage the classifier to rely on voxel patterns that were invariant with respect to the illuminant. Thus, for such ROIs, classification on a new (untrained) illuminant would work.

This was hence the best training scheme to maximize success of our main classification (across illuminants). In order to have a control analysis that was as similar as possible to the main analysis, we used the same training scheme – only the test trials differed between control and main analyses.

2.6.3.2. Across illuminant classification (consistent-cue condition). The critical test for the invariance of surface color representations is to examine how well they generalize across changes in illumination. In other words, how well can a classifier predict surface color when the data used to train the classifier come from different illumination conditions than the data used for testing? Three classifiers were trained to distinguish between blue and yellow surfaces under each two pairs of illuminants, as described above for within-illuminant decoding. However,

testing was carried out on the left-out illumination for each of the classifiers and their accuracies were averaged. In detail: one classifier was trained to distinguish blue and yellow surfaces on responses from “neutral” and “blue-consistent-cue” illumination conditions and tested on responses from the “yellow-consistent-cue” illumination condition; one was trained on data from “neutral” and “yellow-consistent-cue” conditions and tested on data from the “blue-consistent-cue” condition; and one was trained on “blue-consistent-cue” and “yellow-consistent-cue” condition responses and tested on “neutral” condition data. Each classifier was thus trained to discriminate between surface colors under two illuminants and then applied to responses to the same surfaces under the third illuminant that was not included in the training set (Fig. 4b). Note that a change in illumination alters the wavelength compositions of the surfaces and that generalization accuracy above chance indicates that the color representation of the surface is invariant with respect to the illuminant under which it is viewed. It is also important to appreciate that this classification scheme was comparable to the one applied for the classification within illuminants because the same data were used for training.

2.6.3.3. Across illuminant classification (reduced-cue condition). The purpose of the reduced-cue condition was to investigate the sensitivity of any neural signature of constant colors to manipulations that are known to reduce color constancy as measured psychophysically. The classification scheme in this analysis is identical to the one described above with one exception: instead of using the responses elicited under the consistent-cue blue condition and the consistent-cue yellow condition we used the responses elicited by stimuli from the reduced-cue blue and reduced-cue yellow conditions, respectively (Fig. 4b).

2.6.3.4. Ipsilateral and contralateral ROIs. As depicted in Fig. 1a, the four blue or yellow surfaces appeared mostly in one visual hemifield. The surface locations shown in Fig. 1 correspond to the locations for odd-numbered subjects. Their locations were mirrored horizontally for even-numbered subjects. We analyzed the ROI data separately for ipsilateral and contralateral regions with respect to the hemifield where more surfaces appeared.

2.6.3.5. Searchlight analysis. To obtain a more global view of the brain responses coding for surface color, the three types of classifications detailed above were also carried out in searchlight analyses (Kriegeskorte et al., 2006). SVM classifiers were trained and tested in the same way as in the ROI analyses (within illuminants, across illuminants/consistent-cue, and across illuminants/reduced-cue) on local patterns of fMRI responses within a radius of 3 voxels.

2.6.3.6. Statistical inference: ROI results. Non-parametric permutation tests were performed to test whether classification accuracy was above chance (chance level being 50% for binary classifications). Class labels in the training set were shuffled randomly 10^3 times and new classification models were fit to the data and tested on the intact withheld test set. When bootstrapping null distributions for classifications that required averaging across classification accuracies, we took care that the same label permutations were applied in each of the classification analyses involved. This procedure yielded a distribution of mean classification accuracies at the group level that would be expected if no relationship existed between the multivariate data and the class labels in the training set. To correct for multiplicity due to the number of ROIs tested, we controlled the family-wise error (FWE) in the following way (e.g., Blair and Karniski, 1993): within each permutation step, the randomized label assignments were kept identical for all ROIs within individual subjects. Group mean ROI classification values were then obtained for each ROI. Only the maximum group mean value across all ROIs was used for the null distribution. Therefore, a common null distribution was used for all ROIs that controlled the error probability of at least one null hypothesis

being falsely rejected.

2.6.3.7. Statistical inference: searchlight analyses. Individual searchlight maps were spatially smoothed with a 6 mm (FWHM) Gaussian kernel. We tested if decoding was significantly better than chance (i.e., 50%) using a one-sample *t*-test across participants ($df = 18$) and null hypotheses were rejected for *p* values below .001 (uncorrected). The maps were masked with each participant's whole brain mask only.

2.6.3.8. Surface/illuminant bias analysis. The purpose of the aforementioned analyses was to test to what extent surface color representations are invariant with respect to changes in illumination. Since illumination changes alter the SPDs of the light reflected off those surfaces, we considered a complementary question as well: what happens if instead the wavelength distribution of the reflected light is identical for two stimuli that differ with respect to surface and illumination colors? Some brain regions may be biased to encode reflectance, others illumination.

We devised stimuli and performed an analysis that can estimate this bias (Seymour et al., 2015, 2010, 2009): Fig. 8 shows that the light reflected from the yellow surface under blue illumination is identical to the light from the blue surface under the yellow illumination (top-right and bottom-left stimuli in Fig. 8). Classifiers were trained to distinguish between responses to these two stimuli, but tested on responses to the other two stimuli. The classifier results would hence reveal whether a ROI encodes primarily the perceived surface color (i.e. reflectance), or illumination. Either classifier response would be “correct”, but reveal the encoding bias of a given ROI. If the classifier assigns the same labels to stimuli in the same row, the neural representation weights the influence of illumination on the incoming signal more strongly. If it assigns the same labels to stimuli in the same column, the neural representation emphasizes the difference in surface reflectance.

In contrast to our previous ROI analyses, we used a two-tailed permutation technique to test if classifier predictions indicated a representational bias that was significantly different from 50%. This means that the group null distribution this time did not consist of the largest classification value per permutation step but the value with the largest absolute difference from chance level (i.e., 50%). This leads to a bimodal null distribution. Reported *p* values are the proportion of samples in the null distribution that are above or below the observed value (whichever is smaller) multiplied by two.

In order to interpret a bias in favor of illuminant or surface encoding, it is informative to check if illuminant or surface information can be decoded when classifiers are explicitly trained to discriminate along one or the other stimulus dimension. A region that showed no bias for illuminant or surface encoding for instance, would have the same bias as a region that contained only noise. To distinguish between these scenarios, we conducted two control analyses using the same data as in the bias analysis. In explicit surface decoding, classifiers were trained to discriminate between responses to blue and yellow surfaces (under blue or yellow illumination) and tested on how well they could predict surface color in an independent test set (using leave-one-run-out cross-validation). In explicit illuminant decoding, *illuminant* and *surface* switched roles such that classifiers were instead trained and tested on how well they could predict illuminant irrespective of surface color. If activity within a ROI was only noise, none of these classifications would reveal any information.

Finally, we complemented this analysis with a comparison between representational dissimilarity matrices (RDMs) of the measured voxel responses and with two different model RDMs (Nili et al., 2014). For each ROI, we calculated one 4-by-4 RDM. Each entry represented the pairwise dissimilarities ($1 - \text{Pearson correlation coefficient}$) between the average response vectors (voxel values for a given condition of a given ROI) of two conditions. The RDM was calculated for the same four conditions used in the bias analysis. These are given by the two binary factors surface (blue or yellow) and illumination (blue or yellow). The observed

RDMs were tested for agreement with two different model RDMs (shown in Fig. 9c) using rank correlations. The *surface model* assumed that neural activity in a ROI represented differences in surface reflectance but did not distinguish between illuminations. It hence predicted a maximal dissimilarity of 1 between patterns of responses to different *surfaces* and a dissimilarity of 0 between illumination responses (i.e., a complete surface bias). The *illuminant model* made the opposite prediction that neural responses only reflect differences in *illumination* without discriminating between surfaces. It therefore predicted a dissimilarity of 1 between responses to different illuminations and a dissimilarity of 0 between different surface representations (i.e., a complete illuminant bias). Multidimensional scaling (MDS, metric stress criterion) was used to discover relative commonalities between similarity structures in neural encoding within multiple ROIs and the hypothetical model RDMs predicted by theory. We performed hierarchical clustering (average linkage, Euclidean distance) and examined the resulting dendrograms to check whether the representational similarity structures within ROIs preferred a clustering by surface conditions over illumination conditions or vice versa.

2.7. Psychophysics

Human observers differ in their ability to perceive a given surface as the same color when the illumination varies, i.e. in their color constancy. This can be quantified by the color constancy index (CI). We measured color constancy in all our participants behaviorally using an alternating staircase procedure described by Xiao et al. (2012). This method finds the chromaticity of a color that appears achromatic under the illumination of the scenes presented to observers. A CI can be computed with respect to a pair of illuminants from the achromatic settings made under each of them (Brainard, 1998). The color constancy index is based on the notion that a perfectly color constant perceptual system should shift its achromatic point in the direction of the illuminant change while a perceptual system without any color constancy should exhibit identical achromatic points under both illuminants (see [Supplementary Information](#) for formulas). A CI of 0 thus indicates absence of color constancy while a CI of 1 means perfect color constancy (although CIs are not necessarily bound between 0 and 1).

We identified the achromatic point for observers under each viewing condition by instructing them to judge the appearance of a briefly flashed color circle (750 ms) in alternating blocks as either more reddish than greenish or as more bluish than yellowish. The judgments in every trial were used to update the chromaticity of the circle such that it appeared increasingly achromatic. Our methods were identical to those in Xiao et al. (2012) except that we adjusted the chromaticity in the staircase procedure not in RGB space but in a subspace of the perceptually uniform CIE $L^*a^*b^*$ space. To accomplish this, the maximal red, green, blue, and yellow RGB values of the calibrated display were converted to a^* , b^* coordinates. We defined the red/green and the blue/yellow directions as the unit difference vectors between the two coordinate pairs for each direction. Another difference was that the color circle for which the color judgment had to be made did not appear in the center of the screen but in the part of the rectangle where no colored surfaces were present, i.e., in the top-right corner for participants with odd subject numbers, in the top-left for all others.

In addition to determining the CI for each participant, we also determined $CI_{\text{reduced-cue}}$ by applying the same cue conflict manipulation also applied for part of the imaging stimuli (see “*Stimuli: cue conflict conditions*”). This manipulation is known to heavily impair color constancy judgments in human observers (e.g., Delahunt and Brainard, 2004; Xiao et al., 2012), and allowed us to relate its behavioral effect to individual brain decoding results (see section below).

Hence, two CIs were calculated for each participant by taking the average of all CIs in the consistent-cue and reduced-cue conditions, respectively (ignoring conditions in which staircases did not converge, see Results).

2.8. Psychophysical data analysis

The mean CIs in the consistent-cue and reduced-cue conditions were compared using a paired one-tailed *t*-test to ascertain that the cue validity manipulation did in fact show a decrease of color constancy in the cue conflict condition relative to the consistent-cue condition. Finally, we were interested in the psychophysiological relationship between fMRI measurements of neural activity and behavioral measurements of color constancy. Specifically, we tested a prediction from the Equivalent Illuminant Model of color constancy (Brainard and Maloney, 2011). It assumes that color constancy depends on the chromaticity of the illuminant estimated by the perceptual system. Whether color constancy fails or not depends on the accuracy of this estimate (the “equivalent illuminant”). To test this prediction, we calculated the difference in prediction accuracies for the three illuminants between consistent-cue and reduced-cue conditions in each ROI and subject. We then examined if these differences between conditions predict the individual decrease in color constancy induced in the reduced-cue condition using linear regression models.

3. Results

We rendered a complex scene containing four surfaces that appeared either blue or yellow. Three different illuminations were simulated: neutral D65, a blue illumination, and a yellow illumination (Fig. 1a, b, c). Additionally, we introduced a reduced-cue condition for the blue (Fig. 1d) and yellow (Fig. 1e) illumination conditions: in these conditions the background square on which the colored surfaces appeared was replaced with a surface that was chosen such that the light reflected from it was the same as in the neutral condition. This manipulation is known to strongly impair color constancy in human observers (Delahunt and Brainard, 2004; Xiao et al., 2012). We reasoned that a neural correlate of color constancy should be just as susceptible to this manipulation as behavioral color constancy measures. Furthermore, the reflectance of the blue and the yellow surfaces were chosen such that the light reflected from the yellow surface under blue illumination (see power spectra in insets in Fig. 1b on the right) was the same as the light reflected from the blue surface under yellow illumination (see insets in Fig. 1c on the left). This allowed us to study the degree to which different brain regions are biased to encode surface color versus illuminant properties in a situation when the reflected light is physically identical for both stimuli.

Participants completed a neuroimaging session in which they viewed the rendered scenes on a projection screen while lying supine in the scanner and performing a fixation and spatial attention task (Fig. 3). They participated in a behavioral experiment to measure color constancy and a retinotopic mapping experiment.

We trained linear classifiers to discriminate between the two surface colors on the basis of fMRI activation patterns and examined how well they generalize to new samples that were measured either under the same illuminations (Fig. 4a) or a new illumination condition that has not been included in the training set (Fig. 4b). Generalization to a new illumination was used as an indicator for color constancy.

3.1. Psychophysics: cue condition effect

The reduced-cue conditions differed from the consistent-cue ones in that the rectangular background surface of the rendered scene reflected the same light in both blue and yellow illumination: that reflected under neutral illumination. As expected, this manipulation led to a strong decrease in color constancy relative to the consistent-cue condition – in fact, in the reduced-cue condition color constancy was completely abolished (Fig. 5). Accordingly, mean CI was significantly larger in the consistent-cue than in the reduced-cue condition ($CI_{\text{consistent-cue}} = .4716$, $CI_{\text{reduced-cue}} = -.0812$, $t_{17} = 10.566$, $p = 3.4 \times 10^{-9}$, one-tailed). Note that the behavioral data from subject 14 could not be used because the staircases did not converge for any of the illuminant conditions.

3.2. ROI-based fMRI pattern classification: within illuminants

The first fMRI data analysis examined the simplest scenario in which classifiers were trained on samples of the same illuminant conditions as those in the test set. We trained linear classifiers to distinguish between blue and yellow surface colors under pairs of illuminations (e.g. neutral and blue) and tested them on left-out samples from the same illuminations; results were averaged for all combinations of illumination pairs (see Fig. 4a). Training was done for pairs of illuminants to provide a comparison point to analogous subsequent analyses. This analysis does not test for any generalizability across illuminants. It tests for the discriminability of BOLD responses caused by lights of different wavelength compositions, and thus provides a baseline for decoding accuracy for the current data. Our ROIs comprised the visual areas we identified in the retinotopic mapping session and a functionally defined ROI anterior to them found in the searchlight analysis (for ROI definition, see “fMRI pattern classification: searchlight analysis” below). Consistent with previous research (Brouwer and Heeger, 2009), the two colors led to different patterns of fMRI activity in almost all ROIs examined (Fig. 6a), especially in contralateral ROIs. Note that surface stimuli had been presented in both hemifields, but with an asymmetry of three in the contralateral and one in the ipsilateral hemifield (see methods). Classification accuracy ranged from 51.38% (Cohen's $d = .46$) in ipsilateral hV4, marginally failing to reach significance ($p = .057$, FWE corrected), to 54.42% (Cohen's $d = 1.2$, $p = .001$, FWE corrected) in contralateral V1. We conclude that surface color can be predicted from fMRI activity patterns in all ROIs (except ipsilateral hV4) when classifiers have been trained on samples from the same illuminant conditions as those in the test set. Although mean classification accuracies were not much higher than chance, the effect sizes are considerable in magnitude. The result for putative V4α in Fig. 6a is reported for completeness only because it is based on voxels enclosing a cluster of information identified in the same (and therefore non-independent) within-illuminant analysis conducted with the searchlight technique. Note that pV4α results in Fig. 6b and c are independent of the ROI-defining contrast.

3.3. ROI-based fMRI pattern classification: across illuminants (consistent-cue)

This analysis addressed the question whether voxel patterns evoked

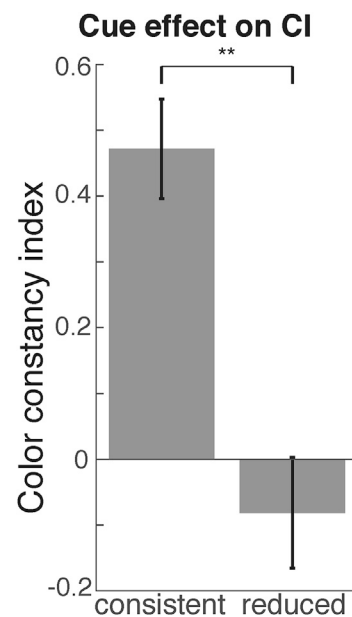


Fig. 5. Behavioral Results. Color constancy was significantly better in the consistent-cue than in the reduced-cue condition (paired *t*-test, $t_{17} = 10.566$, $p < .001$). Error bars denote SEM.

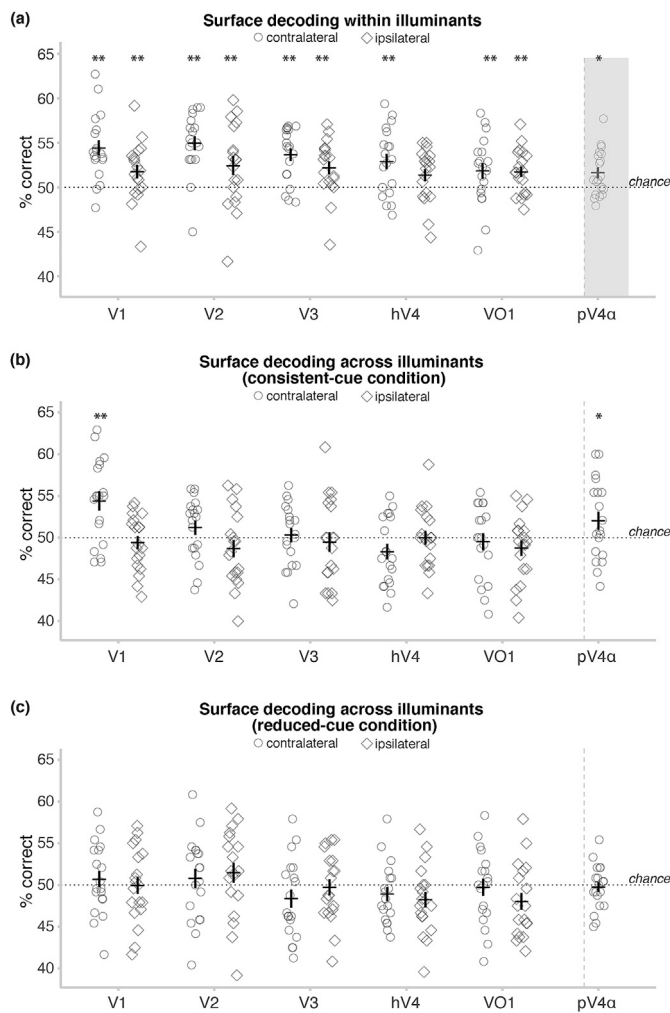


Fig. 6. Pattern Classification Results: ROI Analyses. Mean decoding accuracies in all functionally defined ROIs for surface color discrimination (the three analyses shown in Fig. 3). Results for anatomically defined V5/MT are shown in Fig. S7 in the Supplementary Information. (a) Surface color could be decoded significantly above chance in all ROIs (except ipsilateral hV4) when classification did not involve generalization across illuminant conditions. This panel presents pV4α in light gray as this analysis (within-illuminant decoding) was the same used to define this ROI. Note that the other panels show independent analyses. (b) When classifiers were trained on one set of illuminants and tested on an illuminant not present in the training set, classification was significant only in contralateral V1 and pV4α. (c) In the reduced-cue condition generalization across illuminant condition was not better than chance in any ROI. * $p < .05$, ** $p < .005$, 10^3 permutations, FWE corrected, error bars: SEM.

by the blue and yellow surfaces under one set of illuminants would generalize to new illuminants as well. Generalization of response patterns to new illuminant conditions would imply the invariance of surface color representations with respect to illumination. To this end, we trained classifiers in the same way as in the previous analysis within illuminants but this time tested them on data from an illuminant condition that had not been part of the training set (see Fig. 4b). Our analysis showed (Fig. 6b) that, among the retinotopically defined ROIs, only the V1 area contralateral to the three surfaces allowed predicting the surface color using a classifier trained on responses to stimuli simulating illuminations that were not included in the training set (54.41%, $p = .001$, FWE corrected, Cohen's $d = .88$). Putative V4α also exhibited surface color decoding across illuminants significantly above chance (52.04%, $p = .033$, FWE corrected, Cohen's $d = .43$). We did not, however, observe decoding accuracies significantly above chance in any other ROI, with the highest non-significant decoding accuracy found in contralateral V2 (51.21%, $p = .427$, FWE corrected, Cohen's $d = .33$). Our analysis thus demonstrates that neural representations of surface color in V1 and pV4α

are invariant with respect to illumination changes. This invariance was only observed in these two regions as the patterns of responses to surface color in all other ROIs were not found to generalize across illuminants.

3.3.1. Control analysis: classification across cue conditions

In order to rule out that the null results for across-illuminant surface decoding in hV4 and VO1 was driven by bad signal quality, we performed additional control analyses. In particular, previous research has suggested that fMRI signal quality in V4 can suffer from the presence of nearby blood vessels, which may put this region to a particular disadvantage (Winawer et al., 2010). We trained classifiers to discriminate between surface colors (blue vs. yellow) in the consistent-cue condition (using blocks from the blue and yellow illuminant conditions) and tested them using the corresponding blocks of the reduced-cue condition (and vice versa). The reasoning is the following: the surface-squares in consistent-cue and reduced-cue conditions have identical light emissions (i.e. wavelength-based information). However, the surface *appearances* were more clearly distinguishable as blue and yellow in the consistent-cue compared to reduced-cue blocks. Hence, a region encoding *appearance* (i.e. perception of constant color) should suffer in decoding performance in this cross-testing scenario, whereas a region encoding wavelength-based information should not. Paired t tests in the contralateral hemisphere showed that decoding accuracies were indeed larger in hV4 (56.58%) than in V1 (53.29%, $t_{18} = 2.5617$, $p = 0.0392$, all one-tailed and Bonferroni corrected for four comparisons) and pV4α (50.82%, $t_{18} = 3.9504$, $p = .0019$, Bonferroni corrected, see Fig. S1 for results from all ROIs). For VO1 (53.27%), however, the differences relative to pV4α (50.82%, $t_{18} = 1.4881$, $p = 0.077$, uncorrected) or V1 (53.29%) were not significant. This finding rules out the possibility that signal quality in hV4 was generally worse than in V1 and pV4α.

3.4. ROI-based fMRI pattern classification: across illuminants (reduced-cue)

The observation that classifiers could predict surface color from response patterns in V1 and pV4α even when training and test data came from different illumination conditions raises an important question about the functional role of such activity for color constancy. If it is related to color constancy, there must be an experimental manipulation that simultaneously affects activity in these regions as well as psychophysical measures of color constancy. Our experimental design included a reduced-cue condition, in which a background surface was made to emit the same light in all three illumination conditions. How does this manipulation affect brain activity on the one hand and behavior on the other? As confirmed psychophysically, the reduced-cue condition completely abolished color constancy in our observers (see Fig. 5). The classification analysis was carried out in the same way as in the analysis across illuminations in the consistent-cue condition (i.e., data in training and test set came from different illumination conditions) except that now only responses from the reduced-cue illumination conditions were used (Fig. 1d and e). In this analysis, classification accuracy was not different from chance in any of the examined ROIs (Fig. 6c). Classification accuracy was highest in ipsilateral area V2 (51.47%, $p = .293$, FWE corrected). Importantly, prediction accuracy was not significantly above chance in contralateral V1 either (50.68%, $p = .882$, FWE corrected). Classification of responses in pV4α similarly did not exceed chance (49.76%). Taken together with the psychophysical finding, this analysis shows that a manipulation that strongly impairs color constancy also causes surface color decoding across different illuminant conditions to fail. This suggests that V1 and pV4α activity may have contributed to color constancy in our experiment by encoding surface color in terms of chromatic contrast.

3.5. fMRI pattern classification: searchlight analysis

Since there may be color-responsive activity beyond retinotopically

mapped regions relevant to our task, we repeated the same classifications performed for the ROIs also at the whole brain level by means of the searchlight method (3 voxel radius) (Kriegeskorte et al., 2006).

Fig. 7a shows searchlight results for the within-illuminants analysis. Significant decoding was apparent throughout the occipital cortex including the calcarine gyrus, as well as in the fusiform gyrus primarily contralateral to the surface stimuli. Note that we left-right flipped searchlight maps for participants with even subject numbers to ensure that for all subjects contralateral stimulation was on the right side (positive values of x).

Fig. 7b shows the results from the searchlight analysis across illuminants (consistent-cue). It revealed a cluster of voxels in the fusiform gyrus where differences between local patterns of fMRI activity distinguishing the two surface colors generalized across illuminants. This cluster was located anterior to the ROIs we had examined, with the MNI coordinates of the peak voxel being $x = 34$, $y = -54$, $z = -10$ (Fig. 7b). This fusiform region overlaps with voxels that exhibit classification accuracies above

chance in the within-illuminant searchlight analysis.

As expected from the null-findings in the ROI analyses, also the searchlight analysis on reduced-cue across-illuminant decoding did not reveal any significant results.

The functional properties of the anterior fusiform region thus resemble those of area V1. Both regions allowed predictions of surface colors from local brain activity within the same illuminant conditions as well as across different illuminants. The reduced-cue abolished generalizability across illuminants in both regions. Fig. 7c shows the location of this cluster in relation to the retinotopically defined ROIs in a cortical surface rendering.

Previous studies have already identified two separate color-responsive regions in the fusiform gyrus (Barbur and Spang, 2008; Bartels and Zeki, 2000; Beauchamp et al., 1999; Wade et al., 2008). This region has often been referred to as V4 α . The peak voxel of the cluster in our searchlight analysis was located in close vicinity to the peak voxels listed for V4 α in the review by Bartels and Zeki (2000).

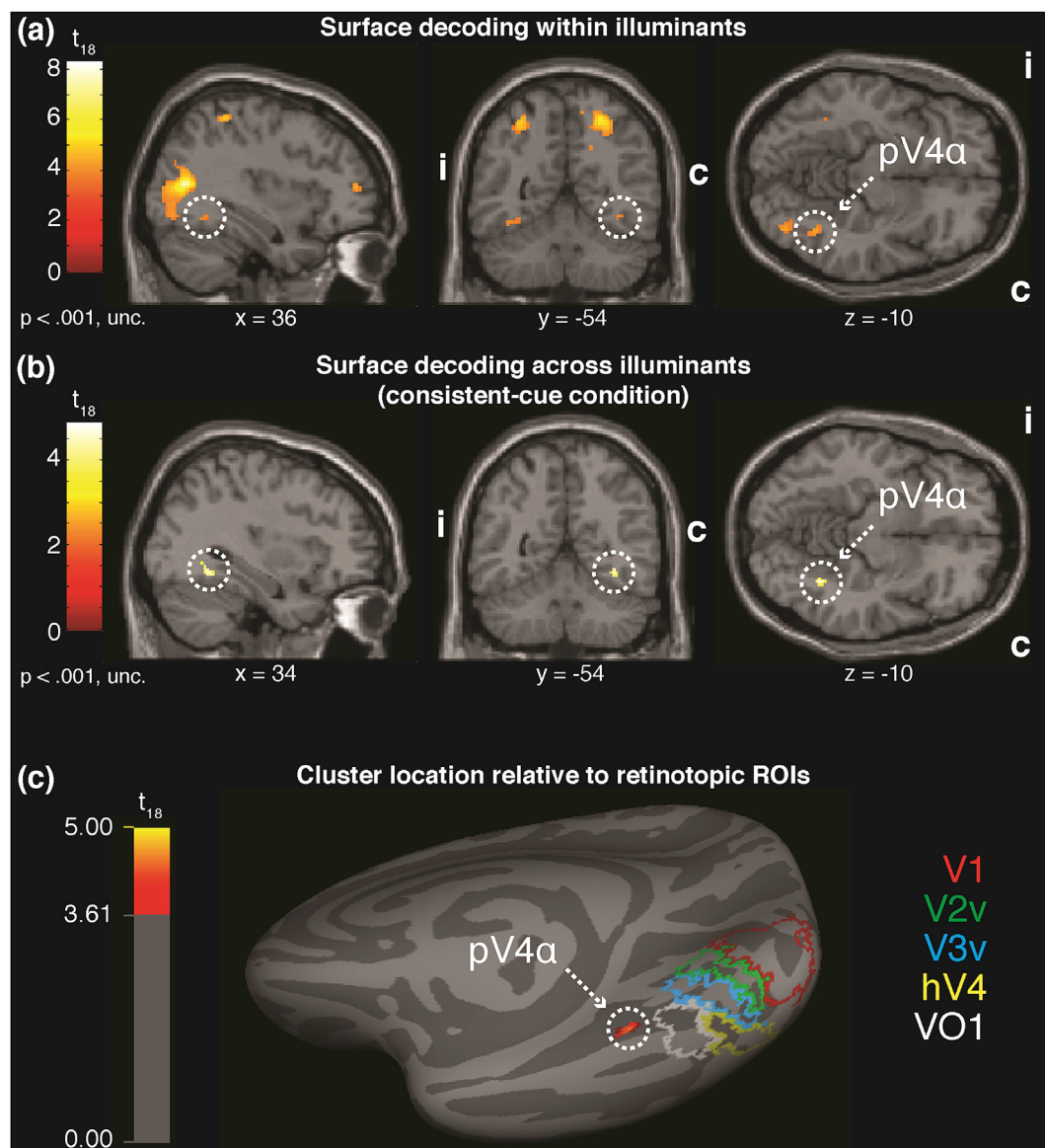


Fig. 7. Pattern Classification Results: Searchlight Analyses. Searchlight maps (3 voxel radius) for the classification analyses shown in Fig. 3 and 5. Letters “i” and “c” denote ipsilateral and contralateral hemispheres with respect to location of most stimuli. (a) Surface classification within illuminants. Circle marks the cluster defined as putative V4 α using within-illuminant decoding. (b) Surface classification across illuminants. The searchlight map reveals a cluster coinciding with pV4 α . (c) Illustration of relative location of the ventral regions. Medial posterior view of the pV4 α cluster identified in the searchlight analysis (dotted circle) and retinotopic ROIs overlaid on a cortical surface rendering in MNI space. The pV4 α cluster was located anteriorly to the ROIs. Colored labels denote surface area falling into individually defined ROIs in at least 25% of the participants.

For a more direct comparison with the results from our ROI analyses, we created a ROI from all voxels within a sphere just large enough to encompass the whole information cluster circled in Fig. 7a (see also “Retinotopic mapping & ROI definition” in Methods and Materials). We included this region as putative V4 α (or pV4 α) in our ROI analyses. Results for this region can be seen to the right of the dotted lines in the plots of Fig. 6.

Due to the similar response properties of V1 and pV4 α , we explored the connectivity between these regions. Partial correlation analyses of the mean residual time series per ROIs showed that there was a small, yet significant amount of unique variance shared between V1 and putative pV4 α ($r = -.038$, $p = .0479$, two-tailed, Holm-Bonferroni corrected for 15 tests, Fig. S2).

3.6. Surface/illuminant bias analysis

It is well known that color constancy depends on our estimation of the illuminant (Delahunt and Brainard, 2004; Xiao et al., 2012). For example, uncertainty about the illuminant within a scene can change the perceived color of a given foreground object dramatically (e.g., “the dress”, Gegenfurtner et al., 2015). It is hence likely that some neural representations primarily encode information about the current illuminant, while others primarily encode the color, i.e. estimated surface reflectance. While our previous analyses focused on the robustness of surface color

representations in the face of changes in illumination and hence wavelength distribution, we here investigated whether ROIs had a bias in encoding surface color or illuminant information when the visual system is presented with lights composed of identical SPDs.

We had designed stimuli such that the yellow surface reflected the same light under blue illumination (SyIb) as the blue surface under yellow illumination (SbIy) (Fig. 8). Hence, while the surfaces emitted identical wavelength information, their colors (i.e. their *perceptual appearance*) differed, as did the context of their illumination. When a classifier was trained to distinguish between responses to these two stimuli, it could rely on two types of information: perceived color (yellow vs. blue surface), or illumination (yellow vs. blue). Which of the two dominated, could then easily be tested by cross-testing this classifier on the two unambiguous stimuli, i.e., blue surface under blue illumination (SbIb) and the yellow surface under yellow illumination (SyIy). If the classifier learned to rely primarily on illumination, it would classify e.g. SbIb as SyIb. However, if it relied primarily on surface color, it would classify SbIb as SbIy.

As can be seen in Fig. 9a, all regions with the exception of pV4 α showed an illuminant bias significantly different from 50% (two-tailed). Areas V1-V3 showed the strongest illuminant bias (weakest in V3: 40.9%, $p = .002$, FWE corrected), followed by areas hV4 and VO1 (weakest in hV4: 46.0%, $p = .01$, FWE corrected). Only pV4 α did not show an illuminant bias (50.9%, $p = .942$, FWE corrected). Since there is no absolute

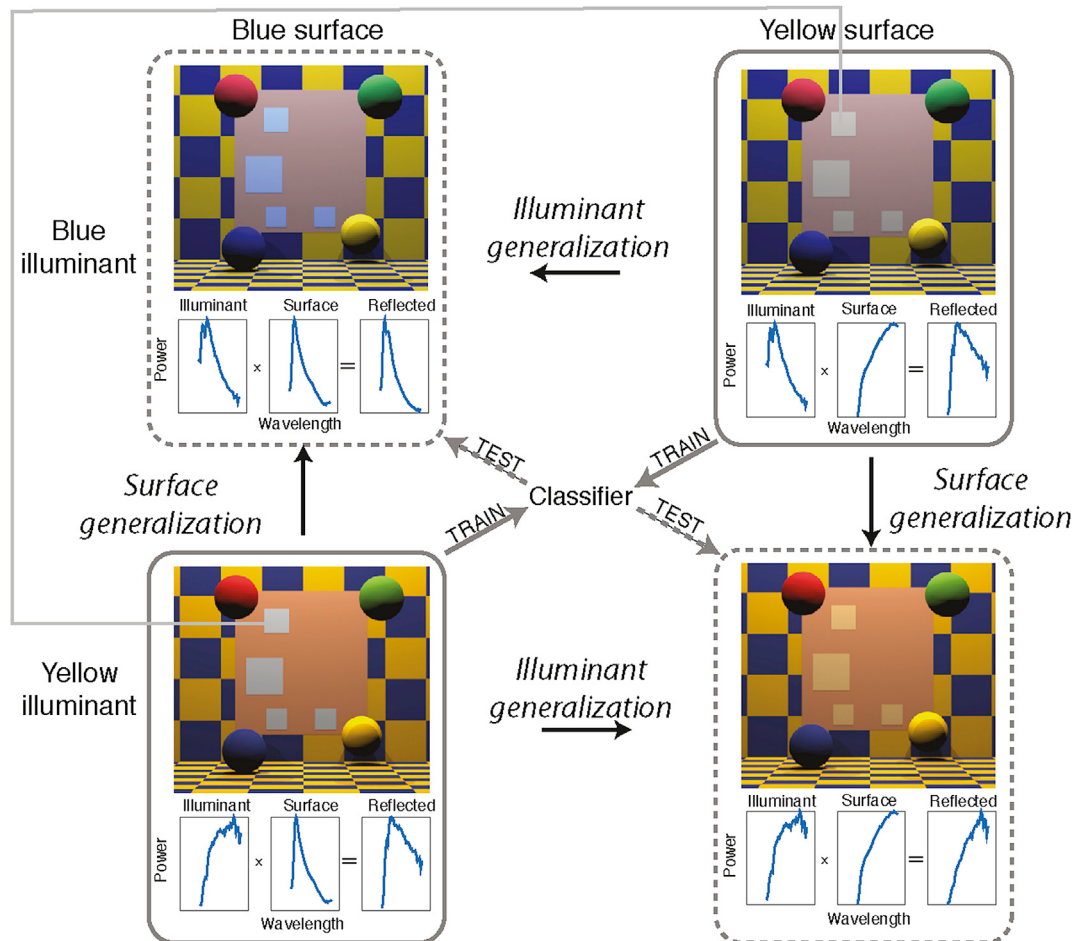


Fig. 8. Surface/Illuminant Bias Analysis: Procedure. Analysis determining bias towards encoding of surface color versus illuminant, respectively. (a) Illustration of stimuli and analysis used for the surface vs. illuminant bias analysis. Stimuli are identical to those shown in Fig. 1b and c. Importantly, the yellow surface under blue illumination reflected the same light as the blue surface under yellow illumination, even though they were *perceived* differently (indicated by gray connector between surfaces). A first analysis trained classifiers to distinguish between these two stimuli. These classifiers could hence rely on *perceptual surface color* or on illumination. Classifiers were tested on responses to the other two stimuli, which revealed which of the two features the classifiers relied on. If stimuli in the same row were assigned the same label, this demonstrated a bias for illuminant representation. If stimuli in the same column were assigned the same label, this indicated a bias for surface representation. In a second and third analysis classifiers were trained and tested to distinguish between illumination (i.e. top versus bottom row) or between surface color (i.e. left vs. right column).

baseline for this bias analysis, we tested for bias differences between ROIs using a repeated measures ANOVA with ROI as fixed and subjects as random factors: differences across all ROIs were significant ($F_{5,90} = 11.126$, $p = 2.51 \times 10^{-7}$, Greenhouse-Geisser corrected for non-sphericity $\epsilon_{GG} = 0.8381$). Post-hoc contrasts with separate error terms

showed that putative V4 α indeed had a stronger surface bias than the remaining ROIs (exceeding Roy-Bose critical value, $F_{5,14} = 5.064$, $p = .0148$, Holm-Bonferroni corrected) and that hV4 and VO1 had stronger surface biases than V1–V3 (exceeding Roy-Bose critical value, $F_{5,14} = 3.149$, $p = .0412$, Holm-Bonferroni corrected).

Since in this analysis 50% could mean “no bias”, but equally well “chance level”, we performed two control analyses using the data from the same four conditions (SbIb, SyIb, SbIy, SyIy). In the first analysis we trained and tested classifiers explicitly using leave-one-run-out cross-validation to distinguish between surfaces (SbIb and SbIy vs. SyIb and SyIy) while in the second analysis discriminations had to be made between illuminants (SbIy and SyIy vs. SbIb and SyIb). As can be seen in Fig. 9b, such classifiers could successfully predict illumination and surface color from almost all ROIs (lowest decoding accuracy for surface decoding in pV4 α : 53.5%, $p = .021$, FWE corrected). Only illuminant decoding was not significantly above chance in this region (51.1%, $p = .653$, FWE corrected). The fact that all ROIs exhibited significant decoding in surface and/or illuminant color decoding shows that none of them represented only noise.

Lastly, we calculated representational dissimilarity matrices (RDMs) to compare the similarity structures reflected in the neural responses within our ROIs with the similarity structures predicted for hypothetical brain regions that exclusively represent illuminant or surface color. While response properties in areas V1–V3 were closest to the illuminant model, pV4 α resembled more the surface model while areas hV4 and VO1 were in between the two (Fig. 9c). Dendrograms on response patterns provide further, descriptive evidence that, while all other ROIs emphasized the difference between illuminants, activity patterns in putative V4 α were clustered primarily according to surface color.

These results demonstrate that there is an increasing gradient from areas V1–V3 to hV4 and VO1 and finally to putative V4 α in preferentially encoding surface as opposed to illuminant color.

3.7. Predicting behavioral color constancy indices from fMRI activity

The Equivalent Illuminant Model proposes a simple two-stage procedure that explains surface color appearance in typical psychophysical experiments: the perceptual system first estimates the chromaticity of the illuminant (the “equivalent illuminant”) and uses this estimate in a second step to calculate surface reflectance (Brainard and Maloney, 2011). Accurate estimates entail better behavioral color constancy. A direct and testable prediction would hence be that the degree to which patterns of fMRI responses to the three different illuminants can be decoded predicts the individual behavioral color constancy score.

We trained classifiers to discriminate between activity patterns elicited by the three different illuminants and cross-validated them by leaving out every run once for testing only. We performed this analysis for the consistent-cue and the reduced-cue conditions, respectively, yielding two decoding accuracies per participant. The decreases in decoding accuracies in each ROI between the two conditions entered simple linear regression models to predict the behavioral decrease in color constancy scores between the two conditions. As Fig. 10 shows, there was a positive correlation in hV4 ($r = .414$, $p = .0436$, uncorrected). The correlation coefficients for the remaining ROIs are shown in Table 1. Although the effect did not survive correction for multiple comparisons, we report the result for hV4 due to its notable effect size.

4. Discussion

The present study is the first, to our knowledge, to investigate fMRI brain responses to surface color that was perceived as constant during illuminant changes, and to relate behavioral color constancy to neural estimates of illumination. We found that the earliest cortical stage, V1, as well as one of the most anterior color-responsive regions in the fusiform gyrus, pV4 α , encode color invariantly with respect to the illuminant. We also found that there is a gradient from early cortex to anterior fusiform

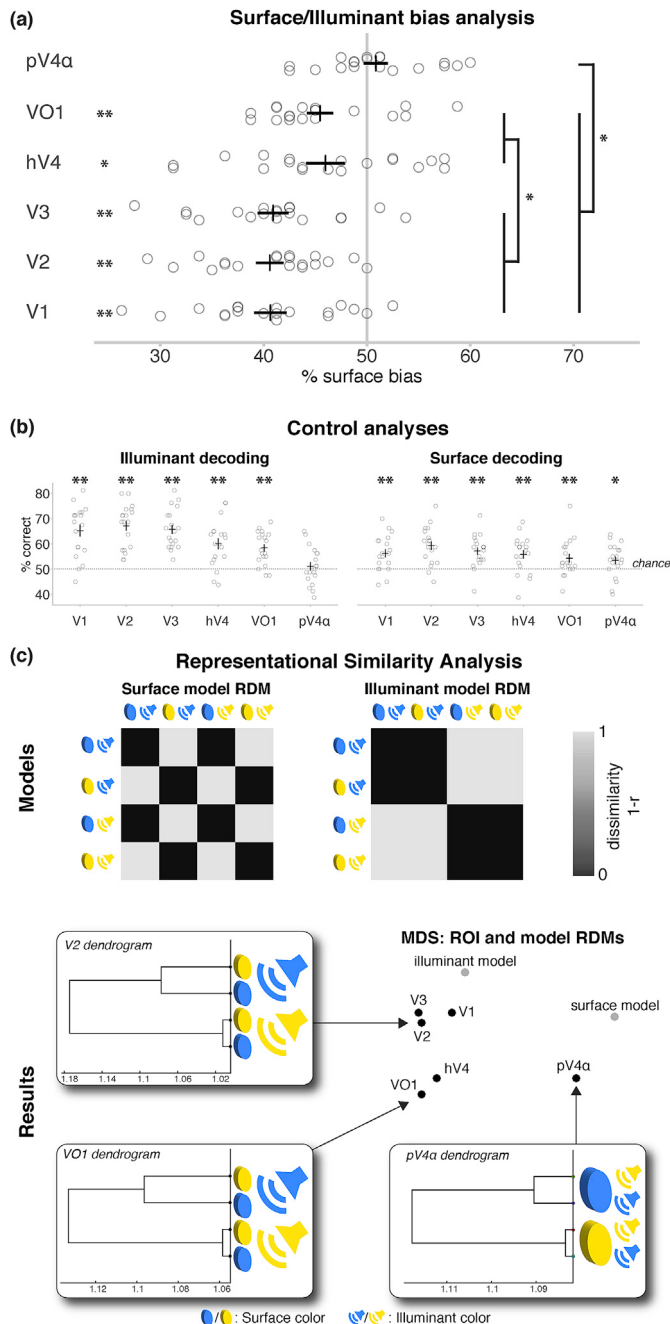


Fig. 9. Surface/Illuminant Bias Analysis: Results. (a) High percentages indicate surface encoding bias while lower percentages indicate illuminant encoding bias. ROI results are tested for a significant deviation from 50% (10^3 permutations, two-tailed). Post-hoc tests showed that pV4 α had a relatively stronger surface bias than retinotopic ROIs and that hV4 and VO1 had stronger surface biases than areas V1–V3. (b) Mean decoding accuracies (error bars represent SEMs) for illuminant and surface color classifications. Accuracies were above chance for both analyses in all ROIs. (c) MDS solution in representational similarity analysis (minimizing metric stress): of all regions examined, areas V1–V3 were most similar to the illuminant model, followed by hV4 and VO1, and lastly pV4 α , which was more similar to the surface model. Dendrograms show that, in contrast to the other ROIs, the grouping of the four stimuli in pV4 α emphasized the difference between surfaces over illuminants.

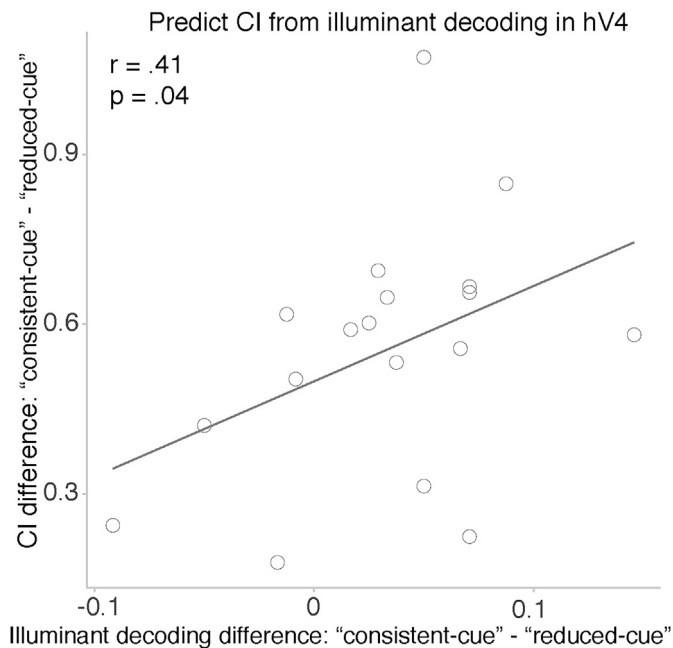


Fig. 10. Predicting Behavioral Color Constancy change due to cue conflict manipulation from neural decoding of illuminant in hV4. Each dot represents a change in color constancy index and the corresponding change in illuminant decoding accuracy between consistent-cue or reduced-cue conditions. Larger differences in illuminant decoding were accompanied by larger differences in color constancy indices.

Table 1

Predicting color constancy from illuminant decoding Correlations between changes in illuminant decoding and color constancy indices in consistent-cue and reduced-cue conditions. Boldface denotes significance ($p < .05$).

ROI	Pearson's r	P (uncorrected)
V1	-.01	.5081
V2	.14	.2946
V3	.06	.4134
hV4	.41	.0436
VO1	.13	.3040
V4 α	-.12	.6853

regions to increasingly encode surface color rather than illuminant. Finally, we demonstrate that illuminant encoding in hV4 predicted the strength of the effect of a cue manipulation on behavioral color constancy, as predicted by the Equivalent Illuminant Model.

4.1. Color constancy computations in V1

We found that surface color could be decoded from fMRI activity in all visual areas when illumination did not change between training and test set. V1 and pV4 α were the only regions where activity encoded surface color such that the information content generalized to new illuminants. The involvement of V1 in color constancy computations fits well with the observation that neurons in this area flexibly adjust their firing behavior to account for chromatic and achromatic changes in the illumination context well outside their receptive fields (RF) (MacEvoy and Paradiso, 2001; Wachtler et al., 2003) and that V1 neurons are spectrally tuned to match the chromatic variability of natural daylight (Lafer-Sousa et al., 2012). Other authors have emphasized the importance of double-opponent cells for color constancy (Conway and Livingstone, 2006; Friedman et al., 2003; Johnson et al., 2001; Shapley and Hawken, 2011): these neurons detect chromatic contrast or color gradients of surfaces, which remain relatively constant across illuminant changes. In keeping with the perceptual relevance of chromatic contrast for surface

perception, a recent study found strong edge enhancement effects for chromatically defined surfaces (Zweig et al., 2015). Similarly, fMRI activity in V1 has been shown to reflect color appearance in perceptual filling-in (Hsieh and Tse, 2010). The two proposed mechanisms are not mutually exclusive and may contribute in a complementary way to color constancy computations in V1 (Hurlbert, 2003). Whatever the underlying mechanism for color constancy in primary visual cortex is, even if it is based on feedback, dysfunction of V1 is known to abolish color constancy judgments in patient D.B. although he could still discriminate between stimuli based on their spectral composition (Kentridge et al., 2007).

4.2. Invariance of surface color representations in V4 across illuminations

We did not find invariant surface color representations in retinotopically mapped areas hV4 or VO1, but more anterior, in pV4 α . These null findings of course do not prove that information is not represented in those areas because differences in decoding accuracies may simply reflect differences in how such information is represented (e.g., chromatic representations in blobs in V1 versus thin stripes in V2), which may in turn influence the “sampling bias” in MVPA (Bartels et al., 2008). Although some differences exist (besides the numerous commonalities) between the physiology of chromatic processing in human and nonhuman primate brains (Lafer-Sousa et al., 2016; Wade et al., 2008), the present findings are in conflict with previous observations that neurons in monkey V4 encode surface color and are robust against changes in wavelength composition (Kusunoki et al., 2006; Zeki, 1983), in particular given the sensitivity of MVPA to surface color in most ROIs in the within-illuminant classification analysis.

Our results may be explained by the fact that neural activity in V4 is strongly influenced by attention to color as demonstrated in both monkey electrophysiology (Maunsell and Treue, 2006; McAdams and Maunsell, 2000; Motter, 1994) and human fMRI studies (Bartels and Zeki, 2000; Brouwer and Heeger, 2013; Saenz et al., 2002). We did not instruct our participants to specifically pay attention to the color of the surfaces in the scene. Alternatively, it is conceivable that V1 and V4 represent chromatic gradients and contrasts at different spatial scales due to differences in RF sizes (for a similar interpretation of Zeki's seminal findings (Zeki, 1983), see Maunsell and Newsome (1987)). Our complete stimulus image ($16.8^\circ \times 15^\circ$), for instance, was considerably smaller than the stimuli used by Kusunoki et al. (2006) (30°). In contrast, Wachtler et al. (2003) examined contextual modulation of chromatic processing at distances of up to only 6° of visual angle from the RF the size of which ranged from 2.5° to 4.5° . In fact, attention may be implemented in form of RF tuning, as has been reported for V4 neurons (David et al., 2008; Klein et al., 2014; Moran and Desimone, 1985). We should point out, however, given that RFs presumably are even larger in V4 α than V4, a pure RF size account cannot fully explain the disparate findings in these two regions.

Our findings are in accord with evidence for the involvement of anterior fusiform gyrus, which includes V4 and V4 α , in achromatopsia (Bouvier and Engel, 2005), with clinical observations being a key reason for the traditional view of this region playing a crucial role in color vision and color constancy (Zeki, 1990). The fact that electrical stimulation in V4 α elicits color percepts in a human patient underscores the relevance of this area for color vision in general (Murphy et al., 2008).

4.3. A potential role of feedback in V1 signal

Finally, given the abundance of cortical feedback to V1 (Felleman and Van Essen, 1991; Muckli and Petro, 2013), it is also conceivable that the information we decode from V1 actually reflects feedback from higher visual areas, possibly V4 or V4 α . Prior studies have shown that BOLD signal is particularly susceptible to feedback (Haynes et al., 2005; O'Connor et al., 2002; Wunderlich et al., 2005), which is most likely due to its strong correlation with postsynaptic neural input (Bartels et al., 2008; Logothetis, 2008). Similarly, numerous fMRI studies that have found decoding or signal modulation specifically in V1, but not in V2,

found this pattern of result in situations where the signal must be due to feedback: for memory color (Bannert and Bartels, 2013), size illusions induced by distance (Sperandio et al., 2012), shape perception (Murray et al., 2002), and context effects in scene perception (Smith and Muckli, 2010). The illuminant invariant surface signals decoded here in V1 may hence also be the result of a complex interplay between V1, hV4 and especially V4 α that encoded illuminant invariant surface color. The significant partial correlation between activity in V1 and V4 α is consistent with this idea (Fig. S2). Similarly, it has been suggested that color could play different roles across areas to either represent illumination or perceived hue (Conway, 2013).

4.4. Functional gradient for surface vs. illuminant color representation from V1-V01

When two differently colored surfaces reflect the same light, they can be discriminated perceptually either on the basis of their reflective properties or based on differences in their illumination. Our surface/illuminant bias analysis tapped into this crucial mechanism of color constancy. The results showed that the tendency to discriminate between illuminant color as opposed to surface color decreased along a gradient from V1 to V4 α (Fig. 9a). The propensity of higher visual areas (in particular hV4, VO1, V4 α relative to earlier areas) to interpret the difference between stimuli as being between surface color (despite matched wavelength composition) is consistent with V4's role in figure-ground segmentation and surface perception (Bouvier et al., 2008; Cox et al., 2013; Poort et al., 2012; Roe et al., 2012) and resembles gradients found for the perception of illusory contours (Mendola et al., 1999) and chromatically defined figure/ground segmentation (Seymour et al., 2015).

4.5. A model for color constancy computations in visual cortex

A remarkable finding in the present study was the correlation of neural discriminability between distinct illuminants with the change in behavioral color constancy indices caused by our cue conflict manipulation. To our knowledge this is the first empirical evidence at the level of neural encoding for the Equivalent Illuminant Model, which links correct illuminant estimation with the ability to estimate surface reflectance (i.e. color) (Brainard and Maloney, 2011). In accord with this, we found that when the difference in neural illuminant decoding between cue conditions was large, the decrease of behavioral color constancy indices was also strong.

5. Conclusion

The present study adds an important new piece to the puzzle of human color vision. Experimental approaches seeking to discover isomorphic mappings between perceptual and neural color spaces have found area V4 to be involved in color perception (Brouwer and Heeger, 2009; Li et al., 2014). In the present study we examined two central components of color constancy in the human brain, namely the robustness of neural encoding of surface reflectance during changes in illumination, and the neural encoding of the illuminant itself. We found that the only regions robustly encoding surface color during varying illumination conditions were the primary visual cortex and a region in anterior ventral cortex previously implied in color vision, pV4 α . Careful stimulus design allowed us to examine for each region whether it was biased in encoding of surface color or the illuminant. This was achieved by choosing distinct pairs of surface reflectance and illumination that resulted in matched reflected light. In such ambiguous situations, there was a gradient from early to higher ventral regions to preferentially encode surface reflectance relative to illumination. Finally, we found evidence suggesting a correlation between perceptual color constancy and neural encoding of the illuminant, as proposed by the equivalent illuminant model, in area hV4.

Acknowledgements

This work was supported by a grant from the German Federal Ministry for Education and Research (BMBF), grant number: FKZ 01GQ1002, by the German Excellence Initiative of the German Research Foundation (DFG) grant number EXC307, and by the Max Planck Society, Germany. The authors declare no competing financial interests.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.neuroimage.2017.06.079>.

References

- Bannert, M.M., Bartels, A., 2013. Decoding the yellow of a gray banana. *Curr. Biol.* 23, 2268–2272. <http://dx.doi.org/10.1016/j.cub.2013.09.016>.
- Barbur, J.L., Spang, K., 2008. Colour constancy and conscious perception of changes of illuminant. *Neuropsychologia* 46, 853–863. <http://dx.doi.org/10.1016/j.neuropsychologia.2007.11.032>.
- Bartels, A., Logothetis, N.K., Moutoussis, K., 2008. fMRI and its interpretations: an illustration on directional selectivity in area V5/MT. *Trends Neurosci.* 31, 444–453. <http://dx.doi.org/10.1016/j.tins.2008.06.004>.
- Bartels, A., Zeki, S., 2000. The architecture of the colour centre in the human visual brain: new results and a review. *Eur. J. Neurosci.* 12, 172–193.
- Beauchamp, M.S., Haxby, J.V., Jennings, J.E., DeYoe, E.A., 1999. An fMRI version of the Farnsworth-Munsell 100-Hue test reveals multiple color-selective areas in human ventral occipitotemporal cortex. *Cereb. Cortex* 9, 257–263.
- Blair, R.C., Karniski, W., 1993. An alternative method for significance testing of waveform difference potentials. *Psychophysiology*. <http://dx.doi.org/10.1111/j.1469-8986.1993.tb02075.x>.
- Bouvier, S.E., Cardinal, K.S., Engel, S.A., 2008. Activity in visual area V4 correlates with surface perception. *J. Vis.* 8 <http://dx.doi.org/10.1167/8.7.28>, 28.1–9.
- Bouvier, S.E., Engel, S.A., 2005. Behavioral deficits and cortical damage loci in cerebral achromatopsia. *Cereb. Cortex* 16, 183–191. <http://dx.doi.org/10.1093/cercor/bhi096>.
- Brainard, D.H., 1998. Color constancy in the nearly natural image. 2. Achromatic loci. *J. Opt. Soc. Am. A. Opt. Image Sci. Vis.* 15, 307–325.
- Brainard, D.H., Maloney, L.T., 2011. Surface color perception and equivalent illumination models. *J. Vis.* 11, 1–18. <http://dx.doi.org/10.1167/11.5.1>.
- Brooks, J., 2012. Counterbalancing for serial order carryover effects in experimental condition orders. *Psychol. Methods* 1–54.
- Brouwer, G.J., Heeger, D.J., 2013. Categorical clustering of the neural representation of color. *J. Neurosci.* 33, 15454–15465. <http://dx.doi.org/10.1523/JNEUROSCI.2472-13.2013>.
- Brouwer, G.J., Heeger, D.J., 2009. Decoding and reconstructing color from responses in human visual cortex. *J. Neurosci.* 29, 13992–14003. <http://dx.doi.org/10.1523/JNEUROSCI.3577-09.2009>.
- Clarke, S., Walsh, V., Schoppa, A., Assal, G., Cowey, A., 1998. Colour constancy impairments in patients with lesions of the prestriate cortex. *Exp. Brain Res.* 123, 154–158.
- Conway, B.R., 2013. Color signals through dorsal and ventral visual pathways. *Vis. Neurosci.* 1–13. <http://dx.doi.org/10.1017/S0952523813000382>.
- Conway, B.R., 2001. Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V-1). *J. Neurosci.* 21, 2768–2783.
- Conway, B.R., Livingstone, M.S., 2006. Spatial and temporal properties of cone signals in alert macaque primary visual cortex. *J. Neurosci.* 26, 10826–10846. <http://dx.doi.org/10.1523/JNEUROSCI.2091-06.2006>.
- Conway, B.R., Tsao, D.Y., 2009. Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex. *Proc. Natl. Acad. Sci.* 106, 18034–18039. <http://dx.doi.org/10.1073/pnas.0810943106>.
- Cox, M.A., Schmid, M.C., Peters, A.J., Saunders, R.C., Leopold, D.A., Maier, A., 2013. Receptive field focus of visual area V4 neurons determines responses to illusory surfaces. *Proc. Natl. Acad. Sci.* 110, 17095–17100. <http://dx.doi.org/10.1073/pnas.1310806110>.
- David, S.V., Hayden, B.Y., Mazer, J.A., Gallant, J.L., 2008. Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. *Neuron* 59, 509–521. <http://dx.doi.org/10.1016/j.neuron.2008.07.001>.
- Delahunt, P.B., Brainard, D.H., 2004. Does human color constancy incorporate the statistical regularity of natural daylight? *J. Vis.* 4, 57–81. <http://dx.doi.org/10.1167/4.2.1>.
- Engel, S., Zhang, X., Wandell, B.A., 1997. Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature* 388, 68–71. <http://dx.doi.org/10.1038/40398>.
- Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., Lin, C.-J., 2008. LIBLINEAR: a library for large linear classification. *J. Mach. Learn. Res.* 9, 1871–1874.
- Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. <http://dx.doi.org/10.1093/cercor/1.1.1>.
- Friedman, H.S., Zhou, H., von der Heydt, R., 2003. The coding of uniform colour figures in monkey visual cortex. *J. Physiol.* 548, 593–613. <http://dx.doi.org/10.1113/jphysiol.2002.033555>.

- Gegenfurtner, K.R., Bloj, M., Toscani, M., 2015. The many colours of 'the dress'. *Curr. Biol.* 25, R543–R544. <http://dx.doi.org/10.1016/j.cub.2015.04.043>.
- Haynes, J.-D., Deichmann, R., Rees, G., 2005. Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. *Nature* 438, 496–499. <http://dx.doi.org/10.1038/nature04169>.
- Heasley, B.S., Cottaris, N.P., Lichtman, D.P., Xiao, B., Brainard, D.H., 2014. RenderToolbox3: MATLAB tools that facilitate physically based stimulus rendering for vision research. *J. Vis.* 14, 1–22. <http://dx.doi.org/10.1167/14.2.6>.
- Hsieh, P.-J., Tse, P.U., 2010. "Brain-reading" of perceived colors reveals a feature mixing mechanism underlying perceptual filling-in in cortical area V1. *Hum. Brain Mapp.* 1407, 1395–1407. <http://dx.doi.org/10.1002/hbm.20946>.
- Hurlbert, A., 2003. Colour vision: primary visual cortex shows its influence. *Curr. Biol.* 13, R270–R272. [http://dx.doi.org/10.1016/S0960-9822\(03\)00198-2](http://dx.doi.org/10.1016/S0960-9822(03)00198-2).
- Ishihara, S., 2011. *Ishihara's Tests for Colour Deficiency, 38 Plates*. Kanehara Trading Inc., Tokyo, Japan.
- Johnson, E.N., Hawken, M.J., Shapley, R., 2008. The orientation selectivity of color-responsive neurons in macaque V1. *J. Neurosci.* 28, 8096–8106. <http://dx.doi.org/10.1523/JNEUROSCI.1404-08.2008>.
- Johnson, E.N., Hawken, M.J., Shapley, R., 2001. The spatial transformation of color in the primary visual cortex of the macaque monkey. *Nat. Neurosci.* 4, 409–416. <http://dx.doi.org/10.1038/86061>.
- Kennard, C., Lawden, M., Morland, A.B., Ruddock, K.H., 1995. Colour identification and colour constancy are impaired in a patient with incomplete achromatopsia associated with prestriate cortical lesions. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 260, 169–175. <http://dx.doi.org/10.1098/rspb.1995.0076>.
- Kentridge, R.W., Heywood, C.A., Weiskrantz, L., 2007. Color contrast processing in human striate cortex. *Proc. Natl. Acad. Sci.* 104, 15129–15131. <http://dx.doi.org/10.1073/pnas.0706603104>.
- Klein, B.P., Harvey, B.M., Dumoulin, S.O., 2014. Attraction of position preference by spatial attention throughout human visual cortex. *Neuron* 84, 227–237. <http://dx.doi.org/10.1016/j.neuron.2014.08.047>.
- Kleiner, M., Brainard, D.H., Pelli, D.G., 2007. What's new in Psychtoolbox-3? *Perception* 36, ECVp Abstr. Suppl.
- Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. *Proc. Natl. Acad. Sci.* 103, 3863–3868. <http://dx.doi.org/10.1073/pnas.0600244103>.
- Kusunoki, M., Moutoussis, K., Zeki, S., 2006. Effect of background colors on the tuning of color-selective cells in monkey area V4. *J. Neurophysiol.* 95, 3047–3059. <http://dx.doi.org/10.1152/jn.00597.2005>.
- Lafer-Sousa, R., Conway, B.R., Kanwisher, N.G., 2016. Color-biased regions of the ventral visual pathway lie between face- and place-selective regions in humans, as in macaques. *J. Neurosci.* 36, 1682–1697. <http://dx.doi.org/10.1523/JNEUROSCI.3164-15.2016>.
- Lafer-Sousa, R., Liu, Y.O., Lafer-Sousa, L., Wiest, M.C., Conway, B.R., 2012. Color tuning in alert macaque V1 assessed with fMRI and single-unit recording shows a bias toward daylight colors. *J. Opt. Soc. Am. A* 29, 657. <http://dx.doi.org/10.1364/JOSA.29.000657>.
- Land, E.H., McCann, J.J., 1971. Lightness and retinex theory. *J. Opt. Soc. Am.* 61, 1–11.
- Li, M., Liu, P., Juusola, M., Tang, S., 2014. Perceptual color map in macaque visual area V4. *J. Neurosci.* 34, 202–217. <http://dx.doi.org/10.1523/JNEUROSCI.4549-12.2014>.
- Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. *Nature* 453, 869–878. <http://dx.doi.org/10.1038/nature06976>.
- Lueck, C.J., Zeki, S., Friston, K.J., Deiber, M.P., Cope, P., Cunningham, V.J., Lammertsma, A.A., Kennard, C., Frackowiak, R.S., 1989. The colour centre in the cerebral cortex of man. *Nature* 340, 386–389. <http://dx.doi.org/10.1038/340386a0>.
- MacEvoy, S.P., Paradiso, M.A., 2001. Lightness constancy in primary visual cortex. *Proc. Natl. Acad. Sci.* 98, 8827–8831. <http://dx.doi.org/10.1073/pnas.161280398>.
- Maloney, L.T., 1999. Physics-based approaches to modeling surface color perception. In: Gegenfurtner, K.R., Sharpe, L.T. (Eds.), *Color Vision: from Genes to Perception*. Cambridge University Press, Cambridge, UK, pp. 387–416.
- Maunsell, J.H.R., Newsome, W.T., 1987. Visual processing in monkey extrastriate cortex. *Annu. Rev. Neurosci.* 10, 363–401. <http://dx.doi.org/10.1146/annurev.neuro.10.1.363>.
- Maunsell, J.H.R., Treue, S., 2006. Feature-based attention in visual cortex. *Trends Neurosci.* 29, 317–322. <http://dx.doi.org/10.1016/j.tins.2006.04.001>.
- McAdams, C.J., Maunsell, J.H.R., 2000. Attention to both space and feature modulates neuronal responses in macaque area V4. *J. Neurophysiol.* 83, 1751–1755.
- Mendola, J.D., Dale, A.M., Fischl, B., Liu, A.K., Tootell, R.B., 1999. The representation of illusory and real contours in human cortical visual areas revealed by functional magnetic resonance imaging. *J. Neurosci.* 19, 8560–8572.
- Mollon, J.D., 1989. "Tho' she kneel'd in that place where they grew..." the uses and origins of primate colour vision. *J. Exp. Biol.* 146, 21–38.
- Moran, J., Desimone, R., 1985. Selective attention gates visual processing in the extrastriate cortex. *Sci. (80-)* 229, 782–784. <http://dx.doi.org/10.1126/science.4023713>.
- Motter, B.C., 1994. Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J. Neurosci.* 14, 2178–2189. doi:8158264.
- Muckli, L., Petro, L.S., 2013. Network interactions: non-geniculate input to V1. *Curr. Opin. Neurobiol.* 23, 195–201. <http://dx.doi.org/10.1016/j.conb.2013.01.020>.
- Murphy, D.K., Yoshor, D., Beauchamp, M.S., 2008. Perception matches selectivity in the human anterior color center. *Curr. Biol.* 18, 216–220. <http://dx.doi.org/10.1016/j.cub.2008.01.013>.
- Murray, S.O., Kersten, D., Olshausen, B.A., Schrater, P., Woods, D.L., 2002. Shape perception reduces activity in human primary visual cortex. *Proc. Natl. Acad. Sci.* 99, 15164–15169. <http://dx.doi.org/10.1073/pnas.192579399>.
- Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., Kriegeskorte, N., 2014. A toolbox for representational similarity analysis. *PLoS Comput. Biol.* 10, e1003553. <http://dx.doi.org/10.1371/journal.pcbi.1003553>.
- O'Connor, D.H., Fukui, M.M., Pinsk, M.A., Kastner, S., 2002. Attention modulates responses in the human lateral geniculate nucleus. *Nat. Neurosci.* 5, 1203–1209. <http://dx.doi.org/10.1038/nn957>.
- Poort, J., Raudies, F., Wannig, A., Lamme, V.A.F., Neumann, H., Roelfsema, P.R., 2012. The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron* 75, 143–156. <http://dx.doi.org/10.1016/j.neuron.2012.04.032>.
- Radonjic, A., Cottaris, N.P., Brainard, D.H., 2015. Color constancy supports cross-illumination color selection. *J. Vis.* 15, 1–19. <http://dx.doi.org/10.1167/15.6.13>.
- Roe, A.W., Chelazzi, L., Connor, C.E., Conway, B.R., Fujita, I., Gallant, J.L., Lu, H., Vanduffel, W., 2012. Toward a unified theory of visual area V4. *Neuron* 74, 12–29. <http://dx.doi.org/10.1016/j.neuron.2012.03.011>.
- Rüttiger, L., Braun, D.L., Gegenfurtner, K.R., Petersen, D., Schönle, P., Sharpe, L.T., 1999. Selective color constancy deficits after circumscribed unilateral brain lesions. *J. Neurosci.* 19, 3094–3106.
- Saenz, M., Burras, G.T., Boynton, G.M., 2002. Global effects of feature-based attention in human visual cortex. *Nat. Neurosci.* 5, 631–632. <http://dx.doi.org/10.1038/nn876>.
- Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Sci. (80-)* 268, 889–893.
- Seymour, K.J., Clifford, C.W.G., Logothetis, N.K., Bartels, A., 2010. Coding and binding of color and form in visual cortex. *Cereb. Cortex* 20, 1946–1954. <http://dx.doi.org/10.1093/cercor/bhp265>.
- Seymour, K.J., Clifford, C.W.G., Logothetis, N.K., Bartels, A., 2009. The coding of color, motion, and their conjunction in the human visual cortex. *Curr. Biol.* 19, 177–183. <http://dx.doi.org/10.1016/j.cub.2008.12.050>.
- Seymour, K.J., Williams, M.A., Rich, A.N., 2015. The representation of color across the human visual cortex: distinguishing chromatic signals contributing to object form versus surface color. *Cereb. Cortex* 1–9. <http://dx.doi.org/10.1093/cercor/bhv021>.
- Shapley, R., Hawken, M.J., 2011. Color in the cortex: single- and double-opponent cells. *Vis. Res.* 51, 701–717. <http://dx.doi.org/10.1016/j.visres.2011.02.012>.
- Smith, F.W., Muckli, L., 2010. Nonstimulated early visual areas carry information about surrounding context. *Proc. Natl. Acad. Sci.* 107, 20099–20103. <http://dx.doi.org/10.1073/pnas.1000233107>.
- Sperandio, I., Chouinard, P.A., Goodale, M.A., 2012. Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. *Nat. Neurosci.* 15, 540–542. <http://dx.doi.org/10.1038/nn.3069>.
- Wachtler, T., Sejnowski, T.J., Albright, T.D., 2003. Representation of color stimuli in awake macaque primary visual cortex. *Neuron* 37, 681–691. [http://dx.doi.org/10.1016/S0896-6273\(03\)00035-7](http://dx.doi.org/10.1016/S0896-6273(03)00035-7).
- Wade, A.R., Augath, M., Logothetis, N.K., Wandell, B.A., 2008. fMRI measurements of color in macaque and human. *J. Vis.* 8. <http://dx.doi.org/10.1167/8.10.6>, 6.1–19.
- Wild, H.M., Butler, S.R., Carden, D., Kulikowski, J.J., 1985. Primate cortical area V4 important for colour constancy but not wavelength discrimination. *Nature* 313, 133–135. <http://dx.doi.org/10.1038/313133a0>.
- Winawer, J., Horiguchi, H., Sayres, R.A., Amano, K., Wandell, B.A., 2010. Mapping hV4 and ventral occipital cortex: the venous eclipse. *J. Vis.* 10. <http://dx.doi.org/10.1167/10.5.1>, 1–1.
- Wunderlich, K., Schneider, K.A., Kastner, S., 2005. Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nat. Neurosci.* 8, 1595–1602. <http://dx.doi.org/10.1038/nn1554>.
- Xiao, B., Hurst, B., MacIntyre, L., Brainard, D.H., 2012. The color constancy of three-dimensional objects. *J. Vis.* 12, 1–15. <http://dx.doi.org/10.1167/12.4.6.Introduction>.
- Zeki, S., 1990. A century of cerebral achromatopsia. *Brain* 113, 1721–1777. <http://dx.doi.org/10.1093/brain/113.6.1721>.
- Zeki, S., 1983. Colour coding in the cerebral cortex: the reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience* 9, 741–765.
- Zweig, S., Zurawel, G., Shapley, R., Slovlin, H., 2015. Representation of color surfaces in V1: edge enhancement and unfilled holes. *J. Neurosci.* 35, 12103–12115. <http://dx.doi.org/10.1523/JNEUROSCI.1334-15.2015>.