

Color statistics of objects, and color tuning of object cortex in macaque monkey

Isabelle Rosenthal

Laboratory of Sensorimotor Research,
National Eye Institute, National Institute of Mental Health,
National Institutes of Health, Bethesda, MD, USA

**Sivalogeswaran
Ratnasingam**

Laboratory of Sensorimotor Research,
National Eye Institute, National Institute of Mental Health,
National Institutes of Health, Bethesda, MD, USA

Theodros Haile

Laboratory of Sensorimotor Research,
National Eye Institute, National Institute of Mental Health,
National Institutes of Health, Bethesda, MD, USA

Serena Eastman

Laboratory of Sensorimotor Research,
National Eye Institute, National Institute of Mental Health,
National Institutes of Health, Bethesda, MD, USA

Josh Fuller-Deets

Laboratory of Sensorimotor Research,
National Eye Institute, National Institute of Mental Health,
National Institutes of Health, Bethesda, MD, USA

Bevil R. Conway

Laboratory of Sensorimotor Research,
National Eye Institute, National Institute of Mental Health,
National Institutes of Health, Bethesda, MD, USA



We hypothesized that the parts of scenes identified by human observers as “objects” show distinct color properties from backgrounds, and that the brain uses this information towards object recognition. To test this hypothesis, we examined the color statistics of naturally and artificially colored objects and backgrounds in a database of over 20,000 images annotated with object labels. Objects tended to be warmer colored (L-cone response > M-cone response) and more saturated compared to backgrounds. That the distinguishing chromatic property of objects was defined mostly by the L-M post-receptor mechanism, rather than the S mechanism, is consistent with the idea that trichromatic color vision evolved in response to a selective pressure to identify objects. We also show that classifiers trained using only color information could distinguish animate versus inanimate objects, and at a performance level that was comparable to classification using shape features. **Animate/inanimate is considered a fundamental superordinate category distinction, previously thought to be computed by the brain using only shape information.**

Our results show that color could contribute to animate/inanimate, and likely other, object-category assignments. Finally, color-tuning measured in two macaque monkeys with functional magnetic resonance imaging (fMRI), and confirmed by fMRI-guided microelectrode recording, supports the idea that responsiveness to color reflects the global functional organization of inferior temporal cortex, the brain region implicated in object vision. More strongly in IT than in V1, colors associated with objects elicited higher responses than colors less often associated with objects.

Introduction

Several lines of evidence suggest that color signals are useful for high-level object vision (Conway, 2018; Tanaka, Weiskopf, & Williams, 2001). Color signals promote object recognition and memory (Gegenfurtner & Rieger, 2000); the addition of color as a feature

Citation: Rosenthal, I., Ratnasingam, S., Haile, T., Eastman, S., Fuller-Deets, J., & Conway, B. R. (2018). Color statistics of objects, and color tuning of object cortex in macaque monkey. *Journal of Vision*, 18(11):1, 1–21, <https://doi.org/10.1167/18.11.1>.

<https://doi.org/10.1167/18.11.1>

Received November 30, 2017; published October 1, 2018

ISSN 1534-7362 Copyright 2018 The Authors



improves performance of object-recognition models (Mely, Kim, McGill, Guo, & Serre, 2016); and color indexing is an effective machine-vision method for identifying objects across image views (Funt & Finlayson, 1995; Swain & Ballard, 1991). Moreover, simulations of object-color statistics suggest the existence of systematic differences in the colors of objects, which could be exploited by the visual system (Koenderink, 2010). But despite the apparent importance of color for object perception, the contribution of color toward the neural operations that support object detection and object recognition have received relatively little attention. There are several reasons that may account for this oversight. First, in the visual neuroscience literature, color has often been assumed to be just a low-level stimulus feature: Color is rarely mentioned in computational work on object recognition (DiCarlo & Cox, 2007; Gauthier & Tarr, 2016); stimuli in object-recognition experiments are often achromatic; and influential object-recognition models ignore color (Riesenhuber & Poggio, 1999). Second, the predominant organizing principle of inferior temporal cortex (IT), the large brain region implicated in object vision, is thought not to depend on color properties of objects, but rather on semantic categories (Huth, Nishimoto, Vu, & Gallant, 2012), or on shape features, such as those that enable a distinction of faces and nonfaces (Kanwisher, 2010), or those that distinguish animate and inanimate objects (Caramazza & Shelton, 1998; Kiani, Esteky, Mirpour, & Tanaka, 2007; Kriegeskorte, Mur, Ruff, et al., 2008; Naselaris, Stansbury, & Gallant, 2012; Sha et al., 2015), or object size (Konkle & Caramazza, 2013). Finally, the boost in fMRI response of many parts of IT caused by adding color to a stimulus is modest (Lafer-Sousa, Conway, & Kanwisher, 2016).

Yet color is possibly the best cue for object vision. Object shape varies with viewing angle, and object motion is not typically an object-defining feature. Object colors, meanwhile, are perceived to be largely constant across viewing conditions. But surprisingly it is not known whether objects can be decoded using only color information; prior work on object recognition has assumed that fundamental object categories, such as the distinction between animate and inanimate, cannot be determined using only chromatic cues (Kriegeskorte, Mur, & Bandettini, 2008).

Here, we revisit the importance of color in object recognition, adopting an approach that aims to understand neural mechanisms in a behavioral context by evaluating natural image statistics (Stansbury, Naselaris, & Gallant, 2013). Prior work has assessed the color statistics of natural scenes (Webster & Mollon, 1997). We are interested in a related, but distinct, issue: the color statistics of the parts of scenes that observers identify as “objects”. To address

this question, we quantified the extent to which color distinguishes objects from backgrounds, by analyzing the hue, chroma, and luminance statistics of parts of images that observers decided were objects (“chroma” being the colorfulness or saturation of a surface as evaluated in the context of a similarly illuminated area that looks white). We used the 20,840-image database originally curated by Microsoft for salient objects (Liu, Sun, Zheng, Tang, & Shum, 2007). We asked observers to mask the images to identify the pixels comprising objects and backgrounds (Figure 1A), to label the objects in the images, and to then group the labels according to superordinate categories “animate” and “inanimate”. We analyzed the color statistics of the images based on several criteria: objects compared to backgrounds, animate objects compared to inanimate objects, and natural objects compared to artificial (man-made) objects; we then compared the ability of classifiers trained on color or shape features to perform object categorization tasks. Next, we directly compared neural color responses across macaque visual cortex with the color statistics of objects, and we confirmed the validity of these fMRI measures through fMRI-guided microelectrode recording. We found that the color tuning of IT was biased towards colors more often associated with objects. The results contribute to our overarching hypothesis that color reflects behavioral relevance (Gibson et al., 2017), and is exploited by the brain to give rise to object knowledge (Conway, 2018).

Materials and methods

Object analysis: Microsoft data set

Color statistics of objects were determined using a large database of images originally curated by Microsoft (Microsoft Research Asia, MSRA) (Liu et al., 2007). The images and the masks that observers in our study used to segment the images into objects and backgrounds, together with the object labels that observers applied to the objects, can be found at <https://neicommmons.nei.nih.gov/#/objectcolorstatistics>.

The images were originally obtained off the internet. We do not hold copyright of the images. The MSRA image set obviated the need for us to define an “object”: Human observers at Microsoft decided whether an object was present in an image, on an image-by-image basis. The objects were identified in photographs, not in real scenes. We make no claim about the relationship of the spectral properties of the objects in the photographs and the spectral properties of the objects as they exist in the real world, although the relationship is not arbitrary because color labels

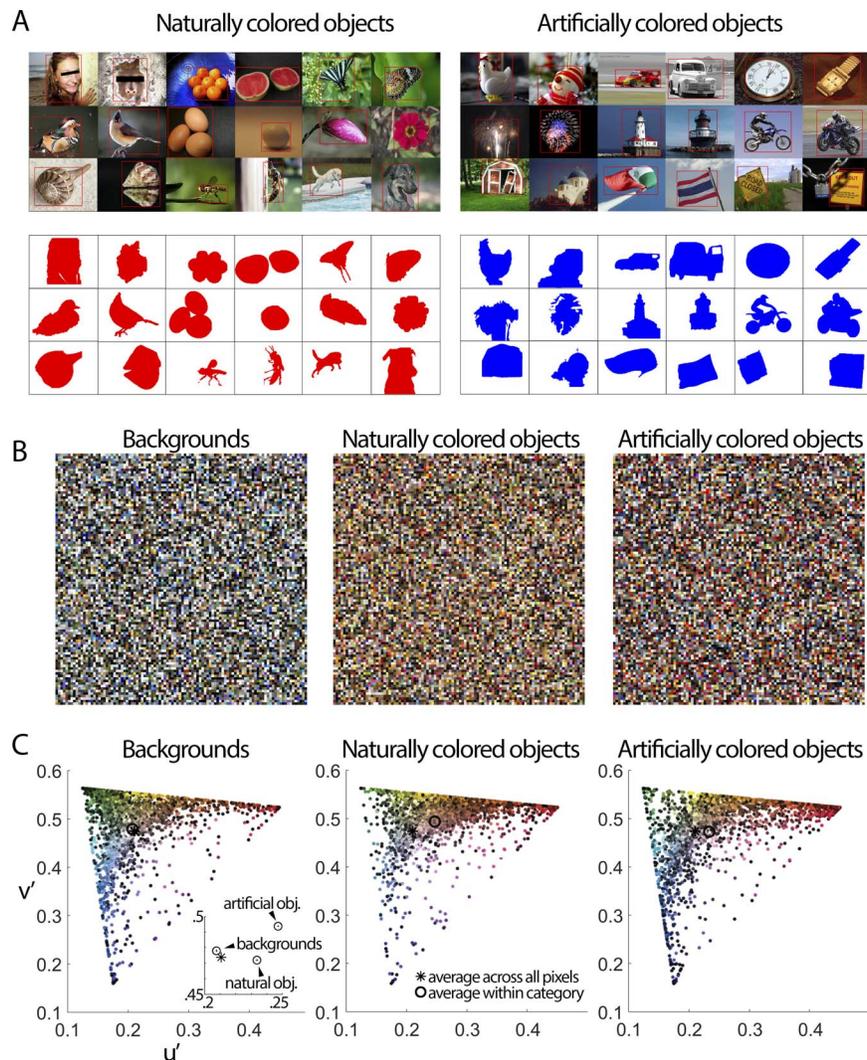


Figure 1. The color statistics of objects. (A) From over 200,000 images, observers at Microsoft selected 20,840 images that contained a salient object and placed a red bounding box around the object (Liu et al., 2007). Two naïve observers in the present study then created masks to demarcate the pixels containing the object. The top panels show example images; the bottom panels show the masks. Images from the MRSA database (Liu et al., 2007) are reproduced with permission from Microsoft. (B) Tapestry plots showing 10,000 pixels randomly selected from the pixels assigned to backgrounds (1.89 billion pixels total), naturally colored objects (0.26 billion pixels), and artificially colored objects (0.22 billion pixels). (C) Chromaticity coordinates of the pixel colors of the tapestry plots. The figure shows a single lightness plane through the chromaticity space, with pixels of all lightness projected onto that plane. The lightness values are indicated by the colors of the data points. Inset shows an expanded view. The symbols contain the standard error of the mean. The analysis was repeated 100 times; each iteration involved randomly sampling 10,000 pixels of backgrounds, artificially colored objects, and naturally colored objects. Pixels from naturally colored objects were different from pixels in backgrounds along the u' dimension ($p = 0$ for all 100 iterations, unpaired t test), and along the v' direction ($p < 10^{-67}$ for all 100 iterations). Pixels from artificially colored objects were different from pixels in backgrounds along the u' dimension ($p < 10^{-82}$ for all 100 iterations), and along the v' direction ($p < 0.05$ for 76/100 iterations). Pixels from artificially colored objects were different from pixels in naturally colored objects along the u' dimension ($p < 10^{-26}$ for all 100 iterations), and along the v' direction ($p < 10^{-130}$ for all 100 iterations).

used to describe objects in color photographs are typically the same as those used to describe the real objects; this conclusion is supported both by empirical work (Moroney & Beretta, 2011), and by the broad acceptance of color photography. People readily, and naturally, identify objects in photo-

graphs; and the use of photographs enables us to assess color statistics of a much larger number of objects than would be feasible if we were to rely solely on spectral measurements from real objects. The database was created by three observers at Microsoft who identified from among 200,000 images a subset of

20,840 images that contained a salient object (Liu et al., 2007). Two people (authors TH and SE) in our group who were initially naïve to the purpose of our study identified the regions within the bounding boxes that contained the object and the background, and identified the object colors as natural or artificial; the regions were identified by loading each image into Adobe Photoshop using a calibrated display and applying a mask. In separate sessions, two observers (author SE and intern JK) subsequently identified and labeled the objects in the image set. A total of 2,781 unique object labels were obtained. Their instruction was simply to label the object so its identity could be communicated to another person. Labeling and applying the masks were done in separate sessions to reduce the possibility that performance on any of the tasks had an impact on performance of the other tasks. The object labels were then categorized as “animate” or “inanimate,” following the procedure of Naselaris et al. (2012). The eyes in Figure 1 have been masked, but original photographs were used in the analysis. A small subset of the photographs (1.7% of the images) was black and white.

The photographs in the image set are not objective random samples of natural scenes, but rather represent photographic decisions. In the same way that studies of “faces” benefit by having a stimulus set with salient examples of faces (even if the stimulus set is not broadly representative of natural images), we reasoned that a study of objects benefits from an image set that is enriched for objects. Humans can identify objects under challenging circumstances such as camouflage and low contrast. But if we want to understand how object recognition works, we think it best to first start with images that unequivocally contain objects that are readily recognized. The objects were not cropped from the backgrounds, so each photograph contains “background” pixels, allowing a direct comparison of the statistics of object and background in each image. The decisions about what constitute “objects” reflects three stages of decision making, by independent observers: the decisions about what to photograph (the photographers); the decisions about what photographs contain objects (the observers in Liu et al); and the decisions about the specific parts of object-containing photographs that comprise the object (the observers in our study). Presumably these decisions reflect the processes of object recognition or object identification, and as such, the photographs allow us to interrogate the statistical basis for object judgements. It will be important to replicate the present results using other image sets, and to probe the extent to which the results are impacted by differences in the way the object-identification task is conducted. These future studies should shed light on how robust the results are

with regards to differences in the instructions used, the background/expertise of the participants, and the conditions under which the object-identification task is performed.

Color spaces used, and why

Colors can be specified in several color spaces that are transforms of each other. CIELUV is traditionally used to specify colors of self-luminous bodies; CIELAB is used to specify colors of reflective surfaces; and cone-opponent spaces are often used in physiological experiments where the goal is to relate color responses to retinal mechanisms of color encoding. CIELUV and CIELAB are very similar, and for consistency, we use CIELUV throughout this report (defining stimuli in the $u' v'$ chromaticity space). The color stimuli used in the physiological experiments were defined using a cone-opponent space (Derrington, Krauskopf, & Lennie, 1984; MacLeod & Boynton, 1979); these stimuli were shown on a computer monitor (a self-luminous body), so we converted these colors into CIELUV to facilitate comparison with the image statistics (see Figure 7B). The conclusions are unaffected if the analyses are performed using CIELAB color space.

Pixel colors from the images (and the color stimuli from the fMRI and microelectrode recording experiments described below) were converted into CIELUV using the image processing and `optprop` toolboxes in MATLAB R2016b (MathWorks, Natick, MA). The images were originally obtained in sRGB values, and these were converted to $u' v'$ coordinates. The white point for the analysis of the image colors was set by D65, the standard illuminant for noon daylight and sRGB color space (Figures 2 and 3), or D50, another standard white point said to represent “horizon light” at sunrise or sunset (all other figures; the results are unaffected by the choice of white point). Pixel chroma was defined as CIE 1976 chroma (MATLAB `rgb2xyz` and `xyz2luvp`, from the `optprop` toolbox), and luminance contrast was computed as Michelson contrast (see Figure 3 legend). Pixel colors were converted into LMS values using custom MATLAB functions with the CIECAM02 transform matrix from XYZ to LMS [0.7328, 0.4296, -0.1624; -0.7036, 1.6975, 0.0061; 0.0030, 0.0136, 0.9834]. Pixel colors were binned into categories defined by equal steps in CIELUV space, or by the stimuli used in the fMRI and neurophysiological recording experiments (each pixel was assigned to the stimulus color with the closest $u' v'$ value, Euclidean distance) using a script written in C++ and OpenCV (`opencv.org`). The fMRI experiments used 12 equi-luminant colors defined by cone-opponent color space (stimulus details given below). The single-cell re-

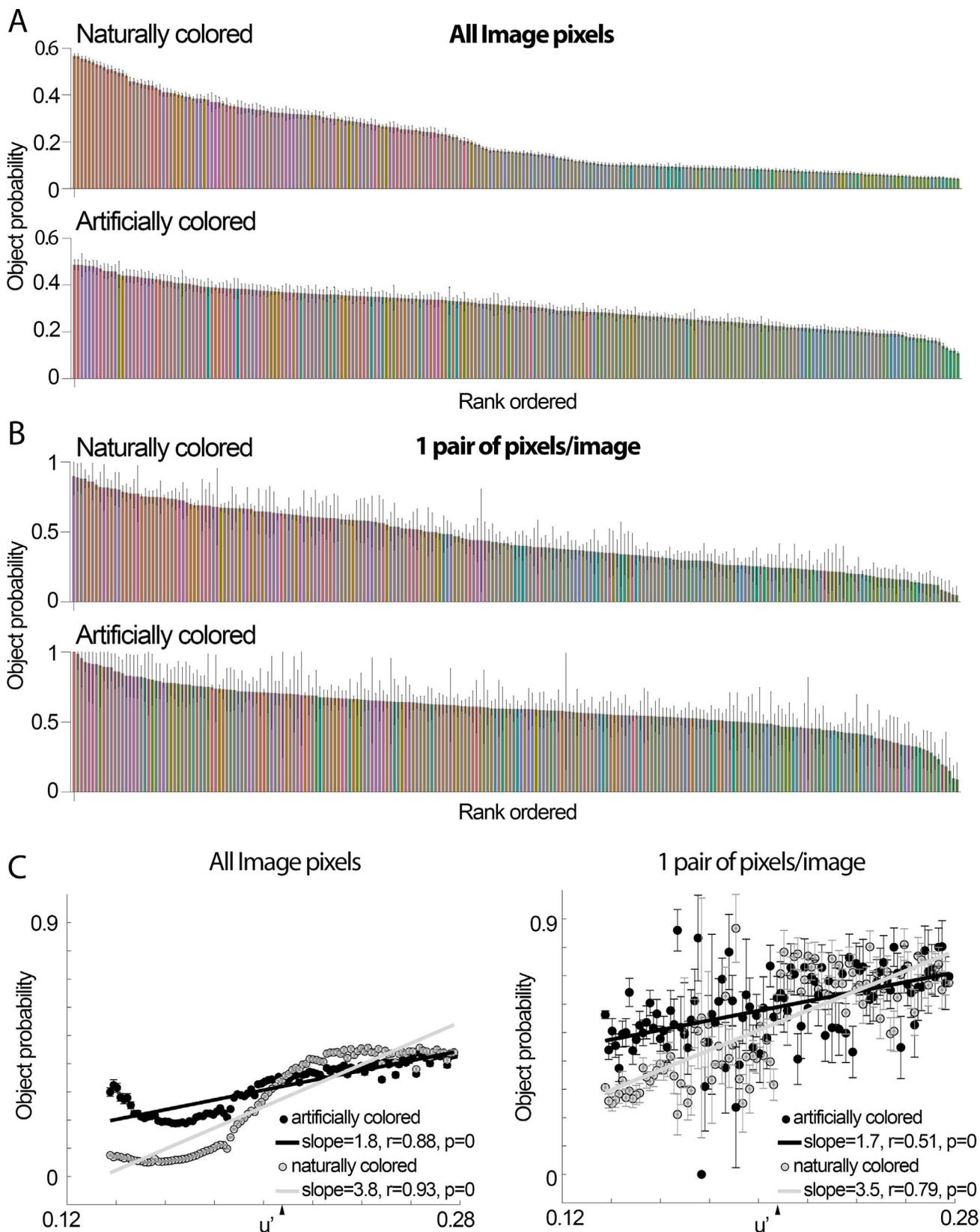


Figure 2. Object-color probability, rank-ordered by most frequent object colors. (A) Pixels across the 20,840 images were categorized into 240 bins (24 evenly spaced hue angles at 10 chroma values in $u' v'$). Object-color probability was computed as follows: number of pixels having a given color in the objects divided by the number of pixels having the same color in the objects + backgrounds. The colors of the bars correspond to the sRGB colors of the bins. The bars are rank-ordered with the most frequently occurring object

→

←

colors on the left. The error bars show the standard deviation of the probability of a given hue being in the foreground, computed over 1000 bootstraps (bar height is the mean of the bootstrapped values). During each bootstrap, 20000 images were picked randomly to generate probabilities. The results across bootstraps were averaged together to generate mean object-background probabilities (bar height); (B) as for panel (A), but using for each image only one randomly selected pixel for each object and one randomly selected pixel for each background. (C) Correlation of object-color probability as a function of u' , for all pixels (left) and single pairs of pixels per image (right). The colors were binned in 101 bins (0.0014 u' bin widths), evenly sampling the u' values symmetric over the white point (arrowhead); error bars show standard deviations, computed as in panel A.

cordings used either 16 equiluminant colors defined by the cone-opponent color space (stimulus details given below), or 45 equiluminant colors that spanned the gamut of the monitor (chromaticity values are given in (Bohon, Hermann, Hansen, & Conway, 2016)). Colors were modulated around a neutral gray, ensuring a constant neutral adaptation state. The gray was $[x, y, Y]$, $[0.3170, 0.423, 55.1]$ for the 12 fMRI colors; $[0.27, 0.296, 55.6]$ for the 16 colors used in the first set of microelectrode recording experiments; and $[0.314, 0.370, 3.05]$ for the 45 colors used in the second set of microelectrode recording experiments. The cone contrast for the 12 fMRI colors is given in Figure 7. The maximum cone contrast for the 16 colors used in the first set of microelectrode recording experiments was ± 0.104 L-M, and ± 0.79 S. The maximum cone activation for the MSRA data set was ± 0.37 L-M and ± 1.63 S.

Object analysis: Color in the Kriegeskorte et al. image set

Support vector machine (SVM) classifiers built using the LIBSVM library (Chang & Lin, 2011) in MATLAB were used to sort the 92 images from Kriegeskorte et al. (2008). Several color statistics were calculated from each image: (a) the average hue of the image averaged over object pixels (we computed the average color across the object, and used the u' and v' values that define this average color to train the classifier); (b) the average luminance of the object; (c) the average color of the image, using L in addition to u' and v' ; and (d) the average chroma of the object. Chroma was calculated as the average Euclidean distance of the pixels in an image from the average pixel value across all images. The SVM classifiers were tasked with discriminating between images labeled as “animate” or “inanimate,” following the classification of the images

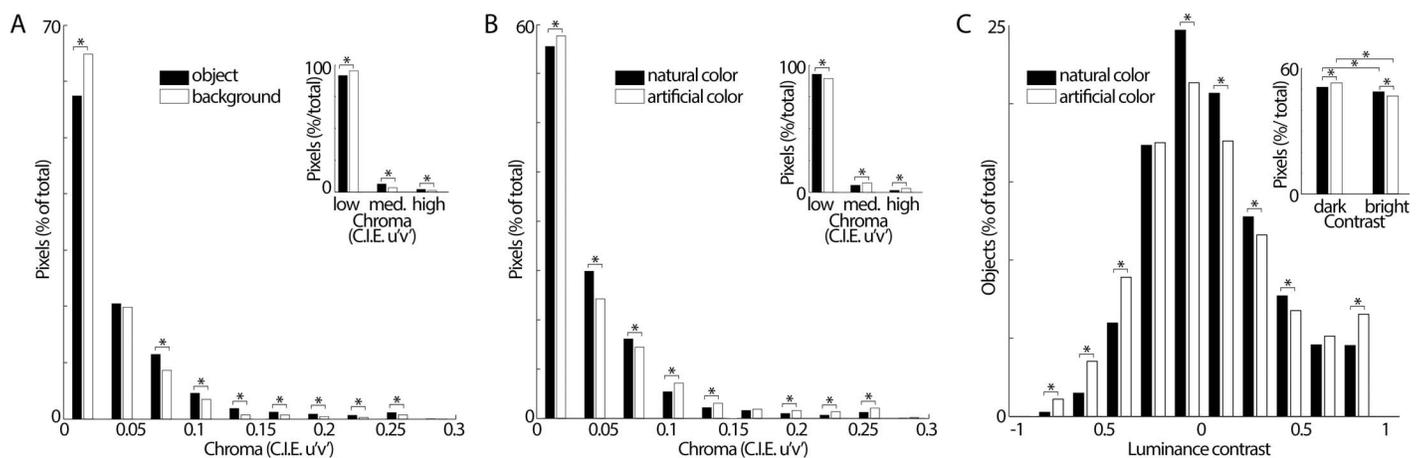


Figure 3. Color chroma and luminance-contrast statistics of objects. (A) Histograms showing the chroma of the pixels identified as part of objects (black bars), and the chroma of the pixels identified as part of backgrounds (open bars; same data sampling as in Figure 2B). Chroma was defined as the CIE 1976 chroma. Insets bin chroma values 0–0.10 as “low,” 0.11–0.20 as “medium,” and 0.01–0.30 as “high”. (B) The chroma of pixels identified as naturally colored (black bars) and artificially colored (open bars). Artificially colored objects were biased towards the chroma extremes, showing more low-chroma pixels and more high-chroma pixels compared to naturally colored objects (this observation is not apparent when chroma values are binned more coarsely, as shown in the inset). Other conventions as in panel (A). (C) The luminance contrast of all image pixels identified as part of naturally colored objects (black bars) are compared against the contrast of the pixels identified as part of artificially colored objects (white bars). Michelson contrast was computed by taking, for each image, the average luminance across the pixels of the object minus the average luminance of the pixels in the background divided by the sum of these luminance values. Insets bin luminance contrast values -1 to 0 as “dark,” and 0 to 1 as “bright”. All differences between pairs of bars identified by an asterisk were significant (chi-square test of proportions, $p < 0.03$).

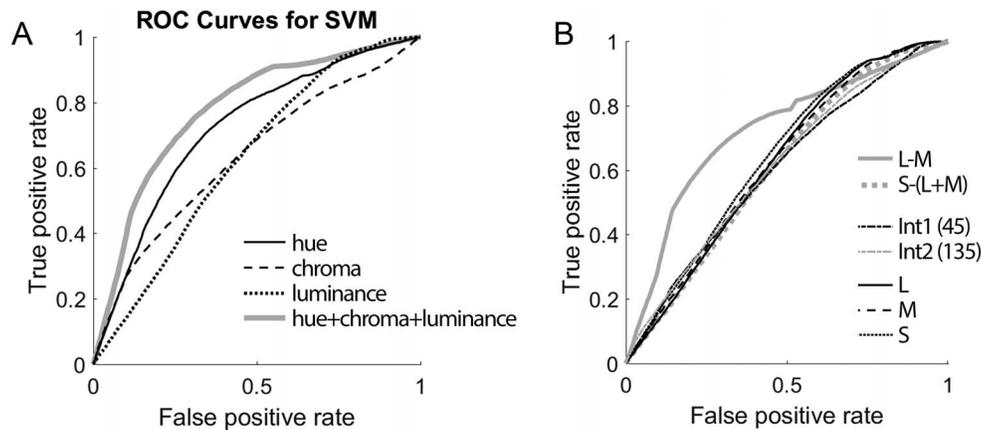


Figure 4. Receiver operating characteristic (ROC) analysis showed that objects can be discriminated from backgrounds based on color. (A) ROC curves when using hue, chroma, and luminance, and all three as input to a support vector machine (SVM) classifier. All ROC curves were significantly different from the null curve (Wilcoxon rank sum test, $p < 0.0001$). (B) ROC curves when using L, M, S, L-M, S-(L-M), and the two intermediate directions in cone-opponent space as features for SVM for classification (Int1 = orange-blue axis; Int2 = green-purple axis). All ROC curves were different from the null curve (Wilcoxon rank sum test, $p < 0.0001$). Only the L-M curve and the S curve were different from all other curves, and each other ($p < 0.0001$).

used by Kriegeskorte et al. For each decoding task, the classifiers were trained on 91 images to generate a model which was tested on the remaining image; this process iterated 92 times, such that each stimulus was the test image in one iteration. Classification accuracy is reported as the proportion of 92 images for which the classifier was accurate and the standard error is of the binomial distribution. We used an SVM classifier with a linear kernel, as well as an SVM with a radial basis function. Linear SVM parameters were selected as follows: for decoding only based on average u' and v' coordinates, cost = 1; $\gamma = 0.01$; for decoding only on luminance, cost = 2; $\gamma = 0.01$; for decoding with L, u' , and v' , cost = 3; $\gamma = 0.01$; for decoding on chroma, cost = 120; $\gamma = 0.01$. Radial basis function decoding parameters for average u' and v' coordinates were: cost = 11; $\gamma = 70$. The cost term reflects a trade-off between correctly categorizing all examples and having the categories separated by a wide margin. Gamma (γ) defines how much influence a single training example has when updating the classifier (larger values meaning less influence). Note that the images used by Kriegeskorte et al. depicted objects that were cropped from natural backgrounds. A broad gridsearch, as recommended by the developers of lib-svm (<https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>), was used to select initial values of cost and gamma, and these values were optimized through a series of narrower gridsearches to optimize performance for each classifier individually. There were 48 animate images and 44 inanimate images in the dataset, giving the classifier a 52% chance of guessing animacy correctly and a 48% chance of guessing inanimacy correctly.

Object analysis: Categorizing object/background of the MSRA image set with color

Support vector machine (SVM) classifiers using different color features were used to classify pixels as belonging to objects or backgrounds. Each image contributed two values: the mean pixel values across the “object,” and the mean pixel values across the “background”. The SVM classifiers trained on half the total data to generate an optimal hyperplane dividing the data into the two classes on the basis of the input features. The trained classifiers were then tested on the other half of the data, returning predictions about class membership for each test image. Each classifier was run 25 times, with a different random 50/50 split of the data for training and testing each time. Accuracy is the mean of these 25 iterations (running more iterations did not cause substantial changes in mean). The standard error of the mean of the binomial distribution was calculated for each iteration, and we report the largest value of the 25 iterations as the *SEM*. Classifiers were run on hue, chroma, and luminance values using D50 as the white point, and with cone values defined in LMS space using the CIECAM02 transformation matrix (Fairchild, 2001). To generate the ROC curves of Figure 4, which quantify the diagnostic ability of the different features, an SVM model was fit to all the image data and object/background labels. MATLAB’s *fitPosterior* function was used to determine the optimal mapping of scores assigned by the classifier to posterior probabilities, and *resubPredict* used this mapping to return the posterior probability of each value being from an object or background. Finally, the *perfcurve* function, using the class labels and the posterior probabilities, generated an ROC curve along with 95% CI obtained through

bootstrapping ($n = 1000$). Additional parameters for all MSRA SVM analyses were as follows: Expected proportion of outliers in training data = 0.05; kernel function = radial basis function; box constraint (regulating how much loss a data point violating the margin incurs) = 1.

Object analysis: Categorizing animate/inanimate of the MSRA image set using shape features versus color features

To compare the performance of color-based features with state of the art computer vision methods in recognizing animate and inanimate objects, we determined the performance of an SVM trained using features generated in three analyses of the MSRA images: first, using color (L u' v' coordinates) alone; second, using a deep convolutional neural network (DCNN; Krizhevsky, Sutskever, & Hinton, 2012); and third, using a histogram of oriented gradients (HOG; Dalal & Triggs, 2005). The early convolutional layers in a DCNN act as Gabor-like filters to capture oriented gradients (Krizhevsky et al., 2012; Zeiler & Fergus, 2014). HOG is based on calculating the image derivative and binning the gradients into a histogram (Dalal & Triggs, 2005). The Gabor-like filters of DCNNs and the image derivative of HOG capture luminance variations in a local neighborhood of pixels, providing a quantitative metric of shape information. Pretrained networks such as the AlexNet have been widely used for transfer learning applications: They are efficient in terms of training, and generally perform well (Huang, Pan, & Lei, 2017; Yosinski, Clune, Bengio, & Lipson, 2014). HOG has also been used in computer vision applications, including distinguishing humans from other objects (Dalal & Triggs, 2005), estimating object pose (Chiu & Fritz, 2015), and general object recognition (Felzenszwalb, Girshick, McAllester, & Ramanan, 2010; Mammeri, Boukerche, Feng, & Wang, 2013). HOG features are reasonably robust to illumination and size (Dalal & Triggs, 2005).

The DCNN that we used is the AlexNet (Krizhevsky et al., 2012), which was trained on the ImageNet (<http://www.image-net.org/>) database. The ImageNet database contains almost 15 million annotated photographs that comprise about 22,000 object categories. Our procedure was as follows: For half of the MSRA image database (about 10,000 photographs), we determined features using the AlexNet DCNN. Then we trained an SVM using the MATLAB implementation (*fitsvm*) on the features from the last fully connected layer of the network and the animate/inanimate labels that we had previously assigned to the images. We tested accuracy of the SVM in assigning animate/inanimate labels on the second half of the image set

	Deep CNN	HOG	CIEL u' v' Color
Animate			
Mean (%)	83.1	54.4	76.0
STD	0.9	0.9	1
Inanimate			
Mean (%)	28.6	73.8	55.0
STD	0.9	1	1

Table 1. Animate versus inanimate classification accuracy.

(images used for training and testing were independent). Each image yields one feature, and the size of the features in the FC8 layer is 1000 dimensions. We performed this procedure 100 times, using a random draw of half of the MSRA image set for each iteration. We trained the SVM using linear, polynomial, and exponential kernels, and used the SVM that obtained the best performance. For the DCNN, the best performance was obtained using the polynomial kernel. Note that the ImageNet database that was used to create the AlexNet DCNN comprises mostly colored photographs, and we generated features using the original colored MSRA images. Thus the DCNN likely exploits both shape and color information.

The analysis using HOG was performed by training the SVM with the features recovered using the HOG implementation in MATLAB. The MSRA images were rendered in black and white, and resized to 300×400 pixels so that the HOG feature vector for each image had the same length. Each image was subdivided into 16×16 cells to calculate the HOG features. We used an SVM with an exponential kernel function to classify the MSRA images into animate and inanimate (the exponential kernel yielded better performance than a linear or polynomial kernel). As above, we used half of the MSRA image set for training the SVM and the other half for testing, and the process was repeated 100 times using a random draw of half of the MSRA image set for each iteration. The HOG procedure used achromatic images, so the features do not reflect a contribution of color.

For the color-alone analysis, we trained the SVM using the three coordinates (L, u', v') obtained by averaging all the pixels within the object boundary of each image as features for that image. We report the results using an exponential kernel for the SVM, which performed better than a linear or polynomial kernel. As in the DCNN and HOG analyses, we trained the SVM using features from half of the MSRA image set, and tested performance on the remaining half of the images. We repeated the procedure 100 times, with a random draw of half of the MSRA images for each iteration.

The mean recognition rate and standard deviation of animate and inanimate classification performance for SVMs trained on the three types of features (color alone, HOG, and DCNN), each for 100 iterations, are

listed in Table 1. The analysis provides a fair comparison because in each case the features (color, DCNN, or HOG) were obtained on the same segmented images. There were a slightly uneven number of animate and inanimate stimuli in the dataset, rendering the odds of a correct “animate” guess at 47% and a correct “inanimate” guess at 53%.

ROC analysis was performed on the color, DCNN, and HOG classifiers in the same manner as described for the object/background SVM classifiers.

Neural measurements of color tuning: fMRI experiments

Color tuning was measured in a Siemens 3-T Tim Trio scanner with AC88 gradient insert, in two alert male macaque monkeys (7–8 kg), using our previously published procedures for MION-contrast-enhanced alert-monkey fMRI (Lafer-Sousa & Conway, 2013; Verhoef, Bohon, & Conway, 2015). Voxel sizes were 1 mm³. The animals were trained to fixate a spot on a screen on which we displayed, in 32-s blocks, each of 12 colors defined by the equiluminant plane of cone-opponent color space (Derrington et al., 1984; MacLeod & Boynton, 1979) (see Figure 7). The colors (Figure 7B) were generated on color-calibrated displays; specification of the color stimuli is given in our prior report (Lafer-Sousa & Conway, 2013). Stimulus blocks were interleaved with 32-s of adapting gray (TR = 2s; echo time = 13 ms). All stimuli maintained an average luminance of 55 cd/m², and were defined to be of equal chroma in CIELUV (note that, typically, stimuli defined in cone-opponent color space are not of equal chroma). The CIE *x*, *y* coordinates for the 12 colors are given in Figure 7. The colors were presented as color-gray gratings (2.9 cycles/°, drifting at 0.75 cycles/s, alternating directions every 2 s) on a screen (41° × 31°) 49 cm in front of the animal using a JVC-DLA projector (1,024 × 768 pixels). In addition to color-gray gratings, we also presented a block of 10% achromatic luminance-contrast gratings. Stimuli were presented in two runs. First run: colors 1, 3, 5, black/white, 7, 9, 11; Second run: colors 2, 4, 6, black/white, 8, 10, 12 (color numbering starts with L-M, and progresses counter clockwise in the cone-opponent color space). To measure color tuning, we analyzed a total of 13,600 TRs (TR = repetition time, 2 s) in one animal and 7,072 TRs in the second animal. The percent signal change (psc) for a given condition was computed against the response to the last eight TRs of the gray conditions before and after. Responses for a given region of interest were computed by averaging the signals across all the voxels within the ROI. High-resolution anatomical scans (0.35 × 0.35 × 0.35 mm³ voxels) were obtained for each animal while it was lightly sedated. Maps of the correlation of

the fMRI response to each color and the probability of each color appearing as part of an object were painted on inflated surfaces of each animal’s anatomical volume. Color-biased regions, face patches, and meridian boundaries demarcating retinotopic visual areas were reproduced from our prior report, using data collected in the same animals (Lafer-Sousa & Conway, 2013; Verhoef et al., 2015). We defined V1 using retinotopic mapping, guided by an anatomical atlas (Paxinos, Huang, & Toga, 2000).

To calculate the correlation between color statistics in the MSRA dataset and the color-tuning responses of voxels in the fMRI dataset (Figure 8B), the pixels labelled as “natural” in the MSRA dataset were binned into the 12 colors used during the fMRI experiment. The probability of the colors in these 12 bins being in objects (rather than background) was then computed. We then determined the correlation of the percent signal change within each voxel within each region of interest with the object-color probabilities, individually for each hemisphere within each monkey. The mean of these correlations across voxels in a unilateral ROI was the average correlation of that unilateral ROI to the object-color probabilities. A paired *t* test was performed to test the hypothesis that the correlation in IT was different than the correlation in V1.

Neural measurements of color tuning: Physiology experiments

Two sets of physiological experiments were analyzed: a first set of experiments in which we targeted color-biased regions of PIT, just anterior to the V4 complex, in the same animals used for the fMRI color-tuning experiments; and a second set of experiments that we have previously described (Conway, Moeller, & Tsao, 2007; Conway & Tsao, 2009), in which micro-electrode recordings were targeted to color-biased globs that are part of the V4 Complex. In the first set of experiments, we measured single-cell responses to 16 colors defined in cone-opponent color space. The CIE *x*, *y* coordinates for the 16 colors were as follows: 0.3129, 0.2784; 0.3341, 0.3296; 0.3503, 0.3949; 0.3547, 0.4633; 0.3382, 0.5077; 0.3026, 0.5013; 0.2636, 0.4478; 0.2355, 0.3781; 0.2213, 0.3155; 0.2177, 0.2684; 0.2218, 0.2367; 0.2304, 0.2172; 0.2421, 0.2089; 0.2567, 0.2102; 0.2730, 0.2208; 0.2922, 0.2430; neutral-adapting gray: 0.2700, 0.2959. In the second set of experiments, we measured single-unit responses to stimuli defined by the gamut of CIE xyY color space (stimulus details are provided in Bohon et al., 2016). The targets in IT and the V4 Complex were identified using fMRI, as previously described (Conway et al., 2007; Conway & Tsao, 2005; Lafer-Sousa & Conway, 2013). Color-biased regions were defined as those showing greater

fMRI activity to equiluminant colored gratings than to achromatic luminance-contrast gratings. In the first set of experiments, the fMRI colors were single color-gray gratings, enabling us not only to identify the color-biased regions but also to assess color tuning across the cortex (details of the fMRI experiments are provided above). In the second set of experiments, the colors used in the fMRI experiment were red-blue gratings of relative luminance contrast defined by a minimum response in area MT. All microelectrode recording protocols were those that we have described previously (Conway, 2001; Conway et al., 2007). Recording chambers were installed over regions of interest, under sterile surgical procedures; custom-made plastic grids fitted to the inside dimensions of the chamber were used to define relative positions within the grid coordinates, and to hold guide tubes through which electrodes were advanced through the cortex. Physiological responses were amplified and filtered, and spikes were sorted either on-line or off-line using Plexon software. While recording, the animals were rewarded with a drop of water or juice for maintaining fixation; eye position was monitored using an infrared camera directed at the eyes. Once a neuron was isolated, bars or patches centered on the receptive field were flashed. The length, width, and orientation of the bar/patch were chosen using hand mapping, to elicit the maximal response. The background of the monitor was a constant, neutral gray, maintaining a consistent adaptation state (7 cd/m^2 in the first experiments; 55 cd/m^2 in the second experiments). The color of the bar/patch was pseudo-randomly selected for each stimulus presentation such that the average number of presentations of all colors was about the same during a recording session. Data were included if we collected responses to at least three presentations of each color. Stimuli were on for 200 ms, and there was a 200 ms delay between stimuli during which time the entire screen (except the small fixation spot) was the neutral-adapting gray. To evaluate the population color-tuning response, and to relate this measurement to the object-color statistics and the color-tuning assessed using fMRI, we determined the average response across the entire population of recorded cells to each of the colors used. This method is different from what we have done previously, in which we conducted population analyses by generating histograms of the number of neurons with peak tuning to each of the colors used (Conway & Tsao, 2009; Stoughton & Conway, 2008). The present method accurately reflects the global number of spikes elicited by each stimulus, and therefore is a more likely cellular correlate of the fMRI response. To relate the microelectrode recording results (using 16 colors) and the fMRI results (using 12 colors), we fit a smooth curve through the color-tuning function of the fMRI data, and subsampled the curve at color intervals defined by

the microelectrode recording stimuli. All experimental procedures conformed to local and US National Institutes of Health guidelines and were approved by the Harvard Medical School, Wellesley College, Massachusetts Institute of Technology, and the National Eye Institute Institutional Animal Care and Use Committees.

Results

Color statistics of objects

The color statistics of natural images are biased for orange and blue (Webster & Mollon, 1997), across ecosystems and seasons (Webster, Mizokami, & Webster, 2007), and are reflected in a subtle bias in V1 for an intermediate color direction that roughly corresponds to the colors associated with the daylight locus (Lafer-Sousa, Liu, Lafer-Sousa, Wiest, & Conway, 2012). Here we extended the approach of relating color statistics to human behavior by testing the extent to which those parts of natural scenes that human observers define as “objects” show distinct color statistics from backgrounds. Our goal is to determine the statistics of those parts of the scenes that humans care about, rather than simply the distribution of colors across natural scenes. We consider this information an important prerequisite towards testing hypotheses about how the brain uses color to generate object knowledge. The Microsoft database (Liu et al., 2007) consists of 20,840 images that contain a salient object; these images were selected from among 200,000 total images. Human observers at Microsoft identified pictures that contained a salient object and then drew bounding boxes around the objects. The database samples many categories of objects, including faces, fruit, butterflies, birds, eggs, flowers, shells, insects, dogs, toys, cars, clocks, fireworks, light houses, motorcyclists, barns, flags, and road signs (Figure 1A). In the present study, two naïve observers segmented the pixels within the object bounding boxes as belonging to objects or backgrounds, and identified the objects as artificially colored or naturally colored (masked regions in Figure 1A).

The colors of the objects were systematically different from the colors of the backgrounds: Backgrounds tended to have cooler colors (bluer, greener, grayer), while objects tended to have warmer colors (yellow, redder), for both naturally colored and artificially colored objects (Figure 1B). Objects were distinguished from backgrounds predominantly in terms of the CIE u' component (Figure 1C, left panel inset), which reflects a strong contribution of the L-M, rather than S, postreceptoral mechanism. These color

statistics are quantified in Figure 1C, plotted in the u' v' chromaticity space, which is designed to be perceptually uniform. The inset in Figure 1C left panel shows an enlarged region of the diagram, showing that the consistent shift in the average color of artificial and natural objects is along the u' direction.

The probability of a given hue being in an object rather than a background correlated with the warmth of the color (the L-M contrast) (Figure 2), especially in the case of natural objects (Figure 2C). Object probability was correlated with u' for both naturally colored objects (Pearson correlation coefficient $r = 0.93$, $p = 0$) and artificially colored objects ($r = 0.88$, $p = 0$) (Figure 2C, left panel). The correlation coefficient was higher for naturally colored objects (Fisher's $z = 1.9$, $p = 0.05$); the average u' value for artificially colored objects was higher than for naturally colored objects (Figure 2C).

One possibility is that the correlations described above are biased because the colors of pixels within an image are not independent. To address this, we randomly sampled a single pair of pixels for each image, one pixel within the “object” pixels and one within the “background” pixels (Figure 2B). This subsampling yielded the same result: Warmer colors had a higher probability of occurring in the “objects” of the images (among naturally colored surfaces, paired t test, $p = 10^{-241}$; among artificially colored surfaces, $p = 10^{-69}$), and the correlation of object-color probability and u' was higher for natural ($r = 0.79$, $p = 0$) compared to artificial objects ($r = 0.51$, $p = 0$; Fisher's z' comparing natural and artificial = 5.8, $p = 10^{-9}$; Figure 2C).

These results uncover three properties of the colors of objects: First, that all objects, regardless of how their coloring is determined (artificial or natural), have a bias for warm colors; second, that the average color among artificially colored objects is warmer than among naturally colored objects; and third, that the correlation between u' (extent of warm coloring) and the likelihood of being an object is less steep for artificially colored objects compared to naturally colored objects. Prior work has shown that objects selected by nonhuman primates in the wild are identified predominantly by differential L versus M signals (Regan et al., 2001). The results in Figure 1 and Figure 2 extend these findings to humans, and to objects that are artificially colored.

The higher probability for objects to have a relatively warm color compared to backgrounds did not depend on chroma level (bars of the same hue but different chroma cluster in Figure 2A and B). Nonetheless, there were subtle systematic differences in the chroma statistics of objects compared to backgrounds. Figure 3 shows the analysis of a single pair of pixels per image. Most pixels across the images (including both objects

and backgrounds) had low chroma (appearing relatively desaturated). But at most chroma levels, objects differed from backgrounds (pairs of bars in Figure 3 identified by an * are different, chi-square test of proportions, $p < 0.03$). Notably, among pixels with highest chroma, more were found in objects rather than backgrounds (inset Figure 3A, right bars, $p = 10^{-12}$), showing that on average objects tended to be more saturated than backgrounds.

The color statistics of naturally colored objects show that among the most saturated pixels in an image, more of them are found in the objects than in the backgrounds. This suggests that high saturation is a feature of objects. If humans are more likely to identify as an object a surface with a relatively high chroma, humans should choose a pigment with relatively high chroma when making objects (on average artificial pigments have a wider gamut than natural pigments). Thus we hypothesized that artificially colored objects would exhibit higher chroma than naturally colored objects. The data support our hypothesis: Among the pixels with highest chroma, more were found among artificially compared to naturally colored objects (inset Figure 3B, right bin, $p = 10^{-12}$). But on close inspection of data binned more finely, the relationship was not that simple: For pixels of the lowest possible chroma, more were artificially colored than naturally colored (main Figure 3B, left-most bin, $p = 0.02$).

We also performed a luminance-contrast calculation. Michelson contrast was computed by taking, for each image, the average luminance across the pixels of the object minus the average luminance of the pixels in the background divided by the sum of these luminance values. Most of the objects of both classes (naturally colored and artificially colored) were of relatively low luminance contrast, with a peak in both distributions that was slightly to the left of zero, indicating a slightly darker contrast on average for all objects (Figure 3C). The results also show that for very high contrasts, objects were more likely to be bright than dark (bars at the right tail of the distribution are higher than bars at the left tail of the distribution). Among the objects of highest contrast (the bins at the tails of the distribution), there were more artificially colored objects than naturally colored objects, paralleling the results on chroma of artificial coloring. When binned into two gross categories of bright or dark, again there were more dark objects than bright objects (naturally colored objects: $p = 0.0013$; artificially colored objects: $p = 10^{-17}$), but there were slightly more naturally colored bright objects than artificially colored bright objects, and slightly more artificially colored dark objects compared to naturally colored dark objects (chi-square test of proportions, $p = 0.005$). Thus artificial coloring is associated with an exaggeration of

the tendency among natural objects for objects to be relatively darker compared to backgrounds.

Classifying objects from backgrounds using color statistics

To quantify the extent to which the brain could exploit the differences in hue, chroma, and luminance of objects compared to backgrounds, SVM classifiers were trained to distinguish between objects and backgrounds. We divided the image set in half and used one half to train the classifier and the other half to test the classifier; we repeated this process 25 times, with different (random) partitions of the data. Using hue (color angle in CIE u' v' space) information alone, the classifier correctly identified an object 64.5% of the time (mean of 25 iterations; standard deviation over the 25 iterations = 1.7%; $SEM = 0.4\%$). We also analyzed a classifier using just chroma (object accuracy = 60.2%, $SD = 1.4\%$; $SEM = 0.3\%$); using just luminance (object accuracy = 48.2%, $SD = 2.1\%$, $SEM = 0.4\%$); and using combined hue, chroma, and luminance (object accuracy = 65.1%, $SD = 2.1\%$, $SEM = 0.4\%$). To quantify how well objects could be distinguished from backgrounds, we ran a classifier on the entire image set using for each image the mean of the pixels comprising the object in the image, and the mean of the pixels in the background. The classifier assigned a posterior probability of being an object for each of these values. The MATLAB function *perfcurve* was used to generate an ROC curve that describes how separable the posterior probability distributions were for objects and backgrounds (the function generated confidence intervals by 1000 bootstraps with replacement; Figure 4A). Classification using all three features (hue, chroma, and luminance) yielded the best performance (area under the ROC = 0.77 [95% CI = 0.77, 0.78]), followed by classification using only the hue feature (area under the ROC = 0.72 [95% CI = 0.71, 0.73]), only chroma (area under the ROC = 0.63 [95% CI = 0.63, 0.64]), and only luminance (area under the ROC = 0.63 [95% CI = 0.63, 0.64]). All ROC curves in Figure 4A were significantly different from the null curve, as assessed by a Wilcoxon rank sum test ($p < 0.0001$). Additionally, the curves were all different from one another ($p < 0.0001$) except one instance: The chroma curve was not different from the luminance curve ($p = 0.08$). These results confirm the observation above that objects are systematically different from backgrounds in terms of their hue.

We next sought to determine for each pixel the extent to which it modulated the three cone classes, using the predicted cone responses as the features for the classifier. The cone responses each in isolation yielded similar classification performance for identifying objects, all hovering around chance ($L = 50.7\%$, $SD = 1.6\%$, $SEM = 0.4\%$; $M = 49.7\%$, $SD = 2.0\%$; $SEM =$

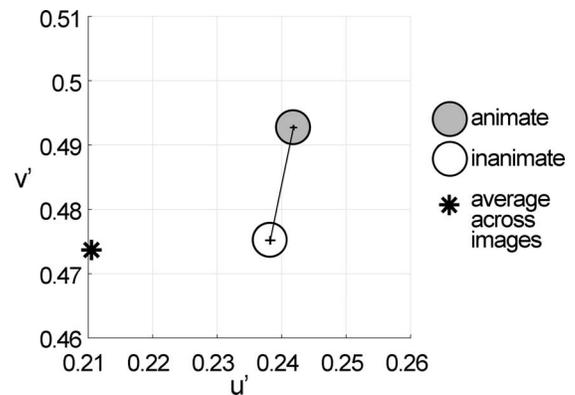


Figure 5. Color statistics differed for animate and inanimate objects. Average chromaticity coordinates for pixels identified as part of animate or inanimate objects. Pixels were drawn from a tapestry of 10,000 randomly selected pixels from each category. This analysis was repeated 100 times (see Figure 2C). Pixels of animate objects were different from pixels of inanimate objects along the u' dimension ($p < 0.05$ for 77/100 iterations), and along the v' direction ($p < 10^{-80}$ for all 100 iterations). Error bars are SEM .

0.4%; $S = 49.0\%$, $SD = 1.8\%$, $SEM = 0.4\%$; Figure 4B). The comparable results for the three cone types are perhaps not surprising because the cone sensitivity functions of the three cone classes broadly overlap. But we found a substantial difference in the classification performance between the two postreceptoral mechanisms: classifiers trained using L-M (object accuracy = 68.1%, $SD = 1.1\%$, $SEM = 0.3\%$; area under the ROC = 0.72 [0.71, 0.73]) performed substantially better than classifiers trained using S-(L+M): object accuracy = 50.9%, $SD = 4.6\%$, $SEM = 0.4\%$ area under the ROC = 0.61 [0.6, 0.61]). This difference is consistent with the observation described above, that the main chromatic feature that distinguishes objects from backgrounds is the extent to which they modulate the so-called “red-green” postreceptoral channel. Classifiers trained using the two intermediate axes in cone-opponent color space were comparable: orange-blue = 47.7%, $SD = 1.6\%$, $SEM = 0.4\%$, area under the ROC curve = 0.63 [0.62, 0.64]; green-purple = 48.7%, $SD = 1.8\%$, $SEM = 0.4\%$; area under the ROC curve = 0.63 [0.62, 0.63]. All ROC curves in Figure 4B were different from the null curve, as assessed by a Wilcoxon rank sum test ($p < 0.0001$). Among the curves, the only two to be significantly different from all other curves on the plot, and each other, were the L-M curve and the S curve ($p < 0.0001$).

Contribution of color to animacy versus inanimacy

The above results show that there are systematic differences in color statistics between objects and

backgrounds that could be exploited by the brain for object vision. Object vision involves distinguishing not only that an object exists, but also the category that it belongs to, and ultimately its identity. Could color provide a signal that informs the computation of behaviorally relevant object categories? To address this question, two observers labeled all the objects in the database of 20,840 images. They subsequently determined whether each object label belonged to the “animate” supercategory, or “inanimate” supercategory. The color statistics of the objects belonging to the animate category were different from those of the objects belonging to the inanimate category (Figure 5A). While the distinction between all kinds of objects and backgrounds involved predominantly a modulation of CIE u' (reflecting the warm bias of object colors shown in Figure 2), the distinction between the two major categories of objects (animate and inanimate) involved predominantly a difference in CIE v' (blue-yellow). These results show that in principle color can provide reliable information for the brain’s object-category computation. We are currently testing whether some specific categories are more distinguishable than others using color, and whether some levels of categorization (basic, superordinate, or subordinate) are more reliably distinguished using color.

Research in monkeys and humans suggests that sensitivity to object animacy is a fundamental operation of the primate visual system, and possibly an organizing principle in IT. When stimuli are grouped by the similarity of the responses they elicit in IT, they form clusters of animate versus inanimate categories (Kriegeskorte, Mur, Ruff, et al., 2008); moreover, the sensitivity to animacy versus inanimacy accounts for the bulk of object-category tuning across IT (Naselaris et al., 2012). As Naselaris et al. (2012) showed, voxels showing stronger tuning to animate objects form a band along the posterior-anterior axis, flanked dorsally and ventrally by voxels preferring inanimate objects. Prior work has argued that color is not a feature used by the brain for animacy calculations (Kriegeskorte et al., 2008). But the results in Figure 5 suggest that color could provide a useful feature for the neural computation of animacy. To test the role of color in bringing about the similarity in responses to items of the same category, we analyzed the 92 images used by Kriegeskorte et al. (2008)—the same images that were used to support the conclusion that animacy categorization is the primary determinant of IT responses. Figure 6A shows the average chromaticity of the images, sorted by animacy. Linear classifiers using several color statistics to identify “animacy” were tested on the images: The most accurate one employed the average chroma of each image as the only feature, at 76.1% correct (SEM of binomial distribution = 4.5%). A classifier that used only luminance (CIE L) was accurate at 56.5% correct

($SEM = 5.2\%$), and a linear classifier using color (CIE L , u' and v') saw no increase in performance (accuracy 56.5%, $SEM = 5.2\%$) suggesting that hue information could not be linearly separated to yield animacy information. Indeed, a linear classifier employing only hue (CIE $u' v'$) was no better than chance at 52.2% correct ($SEM = 5.2\%$). Yet, a $u' v'$ classifier with a radial basis function performed better than any of the linear classifiers at 81.5% correct ($SEM = 4.1\%$). This finding, combined with the high performance of the linear chroma classifier, suggests that the $u' v'$ data of the images can be nonlinearly separated to yield high amounts of information in this dataset about which images are animate, and that this nonlinear information is mostly captured in the chroma of the images. Figure 6B shows that the images formed surprisingly clear animate and inanimate categories when they were sorted with this radial basis function support vector machine, informed solely by the average color of each of the 92 objects.

The SVM applied to the MSRA data, also employing a radial basis function, confirmed that animate and inanimate objects could be distinguished based on color alone. Using only hue angle, the mean classification accuracy for animate objects was 76.0% and for inanimate objects was 55.0% (Table 1), showing that the classifier was better at recognizing animate objects than inanimate objects. The performance of the SVM trained using only shape features (histogram oriented gradient, HOG) or using a deep convolutional neural network (DCNN) that likely reflects contributions of both shape and color were not substantially better. We also computed the area under the ROC for each classifier in Table 1, to better understand how the features each used as inputs were able to provide information about animacy. The Deep CNN AUC was 0.56 [0.56, 0.57]; the HOG AUC was 0.88 [0.88, 0.89]; the color SVM AUC was 0.70 [0.69, 0.71]. A Wilcoxon rank sum test shows that all curves were different from the null curve and from each other ($p < 0.0001$). The classifier results are remarkable: The prevailing view is that shape by itself is necessary and sufficient for categorizing objects. The results here uncover remarkably good categorization of animate and inanimate with hue alone.

Color statistics of objects and color tuning of IT

Most of visual cortex is activated by equiluminant color stimuli (Conway et al., 2007; Conway & Tsao, 2005; Harada et al., 2009; Tootell, Nelissen, Vanduffel, & Orban, 2004), but some regions within the visual cortex show a greater response to equiluminant stimuli than achromatic stimuli. Within the V4 Complex, these color-biased regions, dubbed globs, are on the order of

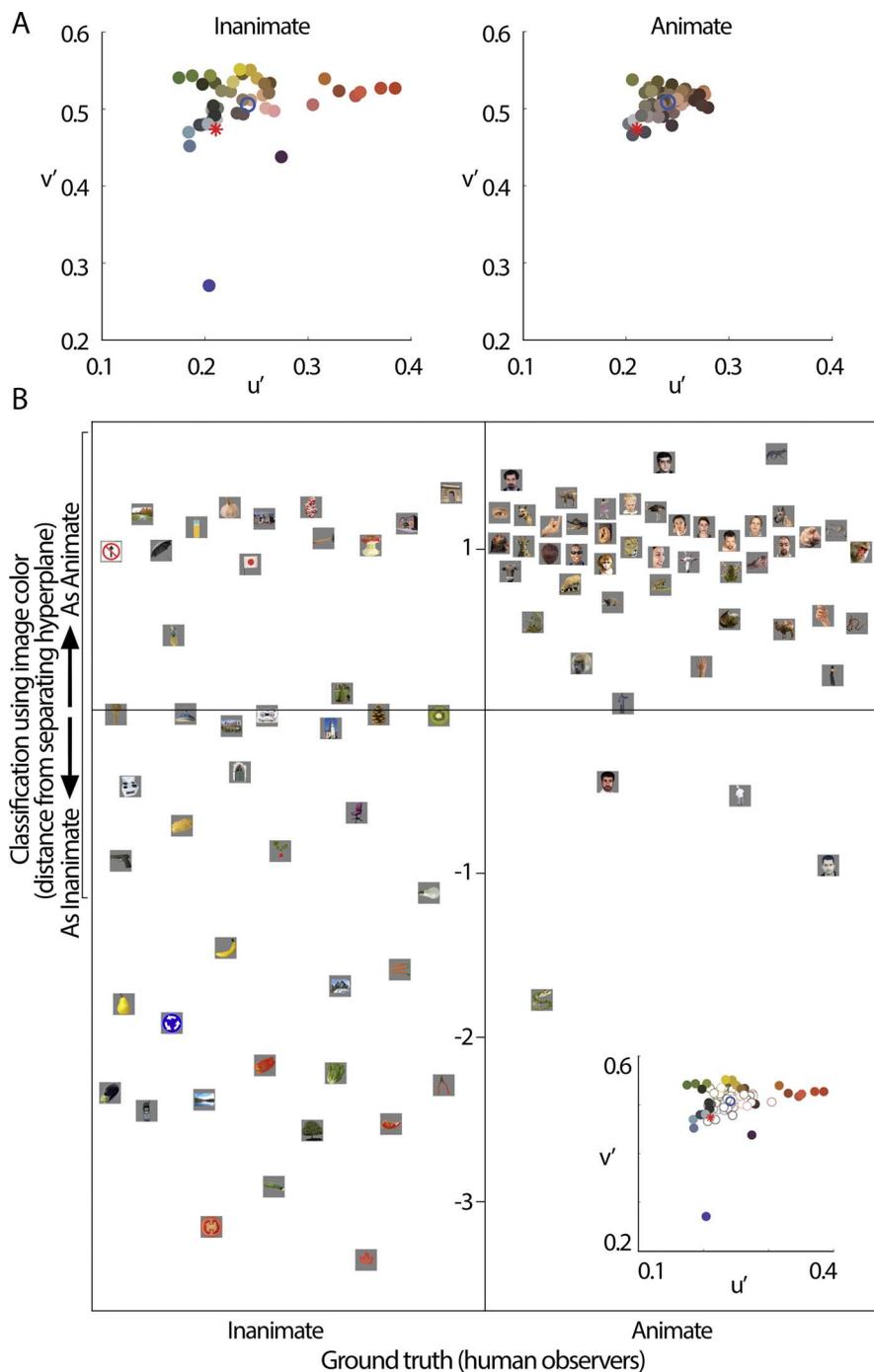


Figure 6. Images could be categorized as animate versus inanimate using only the average color of the object in each image. (A) Chromaticity coordinates for the 92 images used by Kriegeskorte et al. (2008). The asterisks show the neutral point of the color space. The blue circles are the average u' v' value of all the color values plotted in the figure. (B) A support vector machine with a radial basis function trained on only the average CIE u' v' of the object in each image categorized animate images at an accuracy of 91.67% and inanimate images at an accuracy of 70.45%. Inset shows the chromaticity coordinates for all images as in A, in which open symbols indicate images classified as “animate” and filled symbols indicated images classified as “inanimate”. Stimuli are reproduced with permission from Nico Kriegeskorte.

a few millimeters in diameter (assessed with fMRI (Conway et al., 2007) or optical imaging (Tanigawa, Lu, & Roe, 2010)). In IT, color-biased regions are much larger (Conway et al., 2007; Harada et al., 2009),

and are arranged somewhat regularly from posterior to anterior, sandwiched between face-biased regions dorsally and place-biased regions ventrally (Lafer-Sousa & Conway, 2013). Figure 7 shows the color

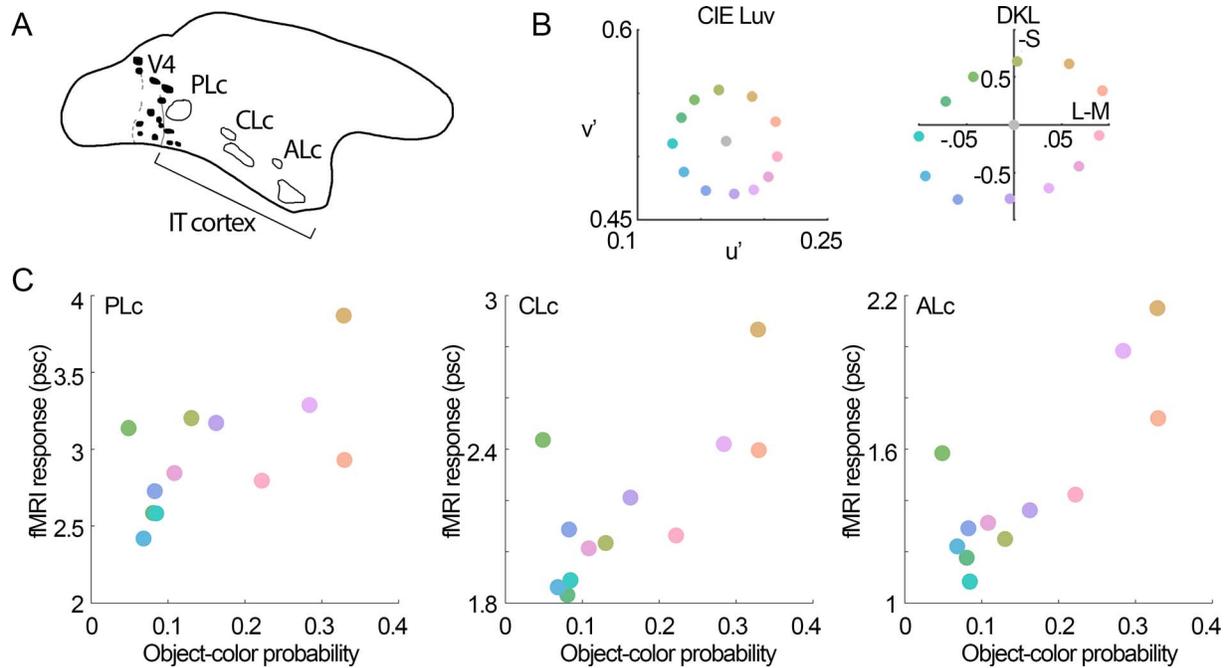


Figure 7. Correlation of object-color statistics and color-tuning of IT color-biased regions. A. Diagram of the lateral view of the macaque brain showing the location of the color-biased regions in inferior temporal cortex (IT), anterior is to the right (Posterior Lateral color, PLc; Central Lateral color, CLc; and Anterior lateral color, ALc; Lafer-Sousa & Conway, 2013). (B) Chromaticities of the colors used in these fMRI experiments. (C) Relationship of the probability of object colors versus color-tuning of color-biased regions within IT. Object-color probability was computed as follows: number of pixels having a given color in the objects divided by the number of pixels having the same color in the objects + background, in natural images. All three regions showed higher responses for colors that were more likely to be the colors of objects (PLc: $r = 0.7$, $p = 0.01$; CLc: $r = 0.72$, $p = 0.01$; ALc: $r = 0.65$, $p = 0.03$). fMRI response (psc) = percent signal change (measured against the signal measured for gray blocks between each colored block).

tuning of color-biased regions identified with fMRI as reported by Lafer-Sousa and Conway (2013), compared with the object-color probability (the object-color probability is the likelihood that a pixel will have a given color given it is part of an object; see Figure 2). All color-biased regions in IT showed greater responses to colors that were more likely associated with objects (Figure 7C).

The results described above prompted us to ask about the color tuning properties across the cerebral cortex. To address this question, for each voxel, we computed the correlation of the color-tuning profile with the object-color statistics (Figure 8A). The strongest correlation with object-color statistics formed a band that spanned color-biased regions and face patches, encompassing parts of IT outside color-biased regions and face patches. V1 showed a much weaker correlation between object-color statistics and color tuning. To quantify this pattern, we computed the average response within V1 and IT of both hemispheres of the two animals tested. The correlations were stronger in IT compared to V1 (Figure 8B; paired t test across V1 and IT, $p = 0.03$).

The pattern of the correlation of object-color statistics and fMRI responses across the cortex suggests

that the neural representation of object colors arises anterior to V1, possibly in V4. As a first test of this hypothesis, we reanalyzed the color-tuning responses from our earlier report (Conway et al., 2007), of the glob and interglob populations within the V4 Complex. The population of glob cells showed a relatively strong correlation with the pattern of object-color statistics ($r = 0.66$, $p = 6 \times 10^{-07}$); interglob cells, on the other hand, showed a lower correlation ($r = 0.30$, $p = 0.04$; significantly different from glob cells: Fisher $z' = 2.2$, $p = 0.02$; Figure 9). More experiments will be required to draw firm conclusions regarding the origin of the cortical bias for object colors, but the present results are consistent with the idea that the linking of objects and their colors is computed within V4/IT.

Discussion

What is an object? There is a strong tendency to think of objects as defined independently from an observer. But in the same way that color is a construct of the brain, what constitutes an object is a manifestation of neural operations responsible for parsing a

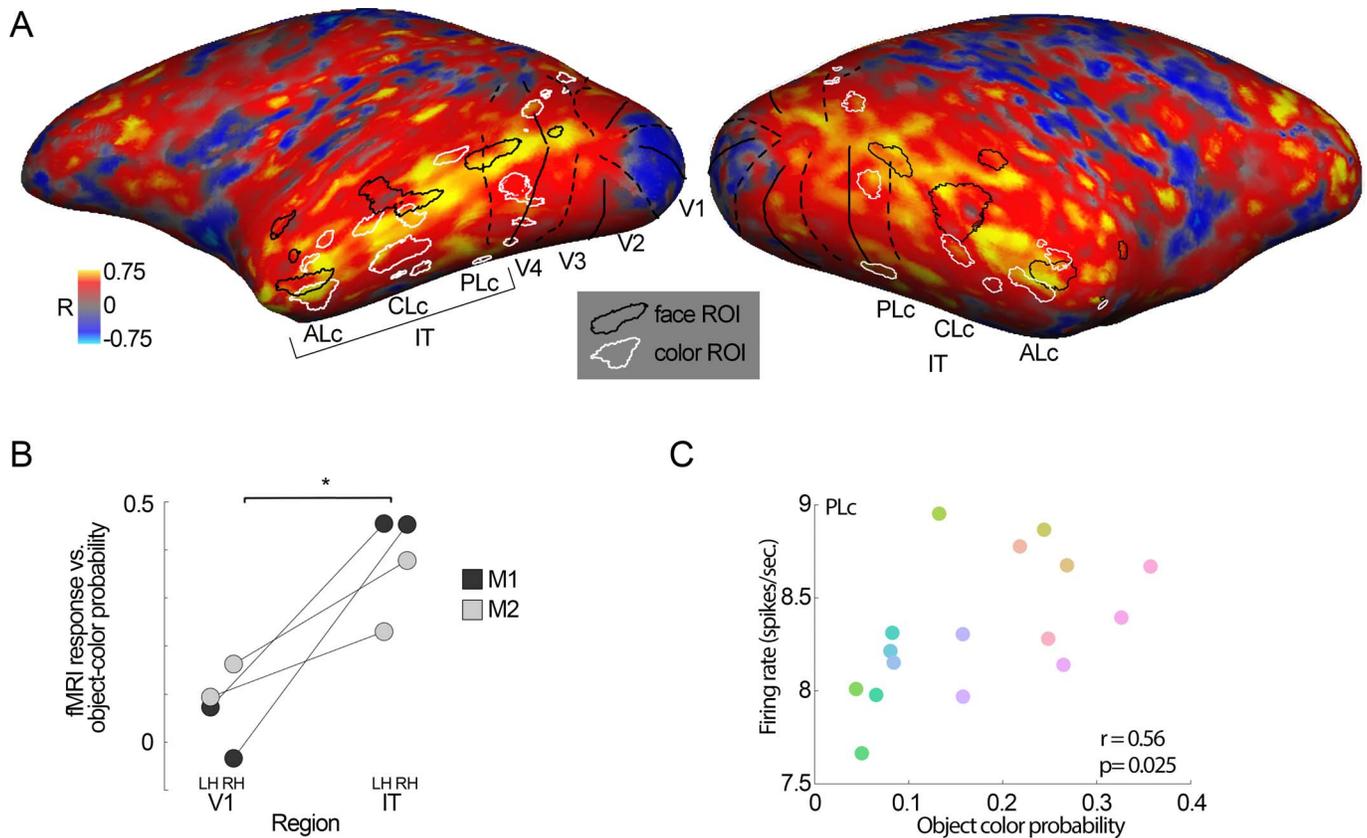


Figure 8. Correlation of object-color statistics across the cerebral cortex. (A) Correlation coefficients between fMRI color tuning and object-color probabilities assessed for each voxel across both hemispheres for one monkey (M1). Color-biased regions are outlined in white; face patches are outlined in black. Vertical meridians are shown as broken lines; horizontal meridians are shown as solid lines. (B) Correlation coefficients were higher in IT compared to V1; the IT region of interest was defined by combining the color-biased regions and the face patches (paired t test across V1 and IT, $p = 0.03$; LH = left hemisphere; RH = right hemisphere). (C) The population of neurons recorded in fMRI-guided recordings of the posterior color-biased region showed a bias for colors more often associated with objects, consistent with the fMRI results ($r = 0.56$, $p = 0.025$). Data were obtained from 117 cells in 54 penetrations in M1.

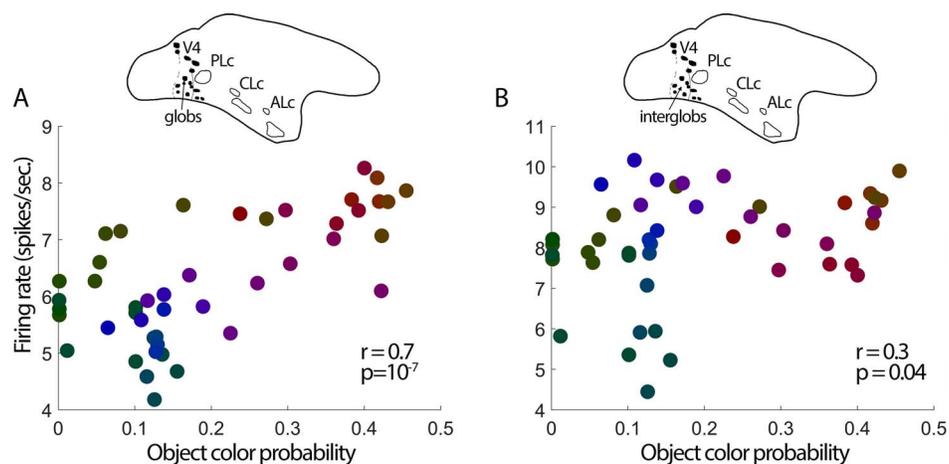


Figure 9. Color-tuning of the population of neurons recorded in posterior IT (the V4 Complex) was correlated with object-color statistics. (A) Responses of cells located within globs (fMRI-identified color-biased regions of the V4 Complex; $N = 300$ cells). The glob-cell population showed a stronger response to the colors more likely associated with naturally colored objects ($r = 0.7$, $p = 10^{-7}$). (B) Responses of cells located within interglob regions ($N = 181$ cells). The interglob-cell population showed a stronger response to the colors more likely associated with objects, but a lower correlation coefficient than was found among glob cells ($r = 0.3$, $p = 0.04$, Fisher $z' = 2.2$, $p = 0.02$).

scene into its useful parts. Prior work has emphasized the importance of shape features in object vision. Here we provide evidence that the brain can also exploit chromatic information to encode objects. We analyzed the color statistics of objects and the relationship of these statistics to neural color tuning. The colors of objects were systematically different from the colors of backgrounds, and the probability of an object having a given color was correlated with color tuning of inferior temporal cortex. We confirmed the correlation using fMRI-guided microelectrode recording of IT and the V4 Complex.

Classifiers trained using color alone not only distinguished objects from backgrounds but also made the superordinate distinction of animate versus inanimate objects, suggesting that color facilitates both object detection and object categorization. The results suggest that brain systems that code hue separately from luminance would be advantageous, providing a functional role for hue representations that are tolerant to variations in luminance (Bohon et al., 2016). The distinction between objects and backgrounds involved predominantly a modulation of CIE u' (reflecting the warm bias of object colors), consistent with the idea that trichromacy evolved to support object vision. The distinction of animate and inanimate objects, meanwhile, involved predominantly a difference in chroma (in both image sets tested) and CIE v' (only in the MRSA image set). The Kriegeskorte image set comprises only 92 images and is perhaps not large enough to uncover the bias in v' found in the larger sample of images.

The coloring of artificial objects might seem arbitrary, but it is not random. The results here show that the color statistics of artificially colored objects reflect, and in some cases accentuate, trends found among naturally colored objects. Natural objects showed a warm bias compared to backgrounds. On average, artificially colored objects showed a higher u' compared to naturally colored objects. Likewise, artificial coloring was associated with exaggerated luminance contrast and saturation of objects. These properties of artificially colored objects were found in the context of a lower correlation, compared to naturally colored objects, between warm coloring and probability of being an object; this relaxation might be attributed to technological developments that have made it possible to create pigments that are relatively rare in nature. Taken together, these results support the idea that color provides an important cue for object vision that is exploited when artificially coloring objects.

The results presented here are the first comparison of the color statistics of objects and neural color tuning, and they provide a foundation for future tests relating neural mechanisms to color cognition. The results depend on the validity of the MSRA image set of

objects. It will be important to replicate the results using other large image sets, and to probe how robust the results are against differences in the object identification task used, the background/expertise of the participants making the object judgments, and the conditions under which the object-identification task is performed. It will also be important to test color tuning in a larger sample of brains (including human brains), as well as neural sensitivity to chroma.

How is color processed in the cortex? Cumulative evidence shows that color processing is not restricted to area V4, or any single “color center”. Instead, color-biased regions in both monkeys (Conway et al., 2007; Harada et al., 2009; Komatsu, 1998) and humans (Bartels & Zeki, 2000; Beauchamp, Haxby, Jennings, & DeYoe, 1999; Lafer-Sousa et al., 2016) stretch across much of visual cortex, including V4 and IT, a swath of tissue implicated in high-level object vision that resides beyond V4 in the putative visual-processing hierarchy. While color-biased regions show especially strong responses to color, most parts of IT that do not show an overt color bias nonetheless respond to equiluminant color stimuli (Lafer-Sousa & Conway, 2013), suggesting that color contributes in diverse ways to the operations of many, if not most, cells in IT. The present observations are consistent with the idea that color contributes to neural computations related to the perception and cognition of objects, and lay the groundwork for future work aimed at determining the specific role of color in color-biased regions and other parts of IT.

Previous work has emphasized the role of IT cortex in computing animacy (see Introduction). The results here suggest that color could contribute to the way in which IT establishes animacy: Classifiers could reliably distinguish animate from inanimate objects using color alone, and IT showed a color-tuning response that was correlated with the object-color statistics. The organizational structure of animate versus inanimate across the cortical surface is correlated with the organizational structure that exists for category-preferring regions: Color-biased regions are sandwiched between face-biased and place-biased regions, forming parallel bands of activity along the posterior-anterior axis of IT (Lafer-Sousa & Conway, 2013). The regions showing strongest preferences for animate objects are centered on the face-processing regions (Naselaris et al., 2012). Our hypothesis—that some part of the animacy response is afforded by systematic variation in color among animate compared to inanimate objects—therefore predicts a systematic difference in the color-tuning within IT, which the electrophysiological results support. The peak correlation of fMRI color tuning and object-color probability formed a band running along the length of IT, centered on the face patches. L-M signals (warm coloring) relays social information

in monkey and human faces (Rhodes et al., 1997; Stephen, Law Smith, Stirrat, & Perrett, 2009); we hypothesize that the warm coloring of socially important stimuli is partially responsible for driving the organization of IT, exerted either during evolution or development.

The potential role of color in the organization and operations of IT may have been overlooked because many investigations of object recognition have used achromatic images. The choice to use achromatic images reflects an implicit understanding of the potential confounds introduced by color, and a desire to control for them. But ironically, one consequence of this decision may be the widespread belief that color is not important for object recognition—a belief that may have led to the uncontrolled use of color in some contemporary research. The consequence is that the organizational patterns in IT associated with animacy that were attributed to semantic (or shape) features, may be caused in part by systematic differences in the colors of animate versus inanimate objects (or systematic differences in natural versus artificial coloring). In any event, the present work underscores the likely importance of color to high-level object processing, possibly beyond the animate-inanimate (or natural-artificial) category distinction. Among the 20,840 images, observers identified 2,781 unique object labels. We are currently testing the extent to which color can classify each of these object types. Determining this information could prove useful in exploiting color for the design of artificial object-recognition systems. One hypothesis is that classification will be better for natural objects of high behavioral relevance.

What is color for? Color is clearly not the only property used in neural computations supporting object vision, nor is it an essential feature for all object-recognition tasks. But behavioral evidence regarding the importance of color to vision (Conway, in press; Gegenfurtner & Rieger, 2000; Rowland & Burriss, 2017; Tanaka et al., 2001), along with the results presented here, behooves the incorporation of color into mainstream object-recognition algorithms (Zhang, Barhomi, & Serre, 2012) and predictive encoding models (Naselaris et al., 2012). Yet how should color be incorporated into these algorithms? Most object-recognition algorithms have failed to find a strong advantage when color edges are incorporated as an additional shape feature, which raises the possibility that the information about objects that color provides is distinct from the information provided by luminance contrast (shape). The history of technology is characterized by an ever-increasing push to develop new ways of making colors (Finlay, 2014), and a desire for a larger gamut, even though this information ostensibly provides little benefit for object recognition (Lee & Ju, 2017; Zamir, Vazquez-Corral, & Bertalmio, 2017).

Instead of suggesting a role of color in object identification, the history of color science underscores the important association of color and subjective experience (e.g., pleasure, emotion): Color does not tell us what an object is, but instead whether we should care about it (Conway, 2018). Consider a banana. Over its lifespan, its shape changes little, while its color changes substantially. The shape tells us what it is; the color tells us whether we want to eat it. In this framework, an important role of color in behavior is to provide valence information about objects. Indeed, color can provide information about material changes that often accompany changes in valence (Yoonessi & Zaidi, 2010). We wonder whether some circuits in IT exploit shape information to determine the identity of objects; and separate networks within IT use color information to establish the extent to which we care about the objects. Formalizing such a model could impact object-recognition efficiency: Rather than processing for recognition everything within the visual field, the system could exploit color information to identify the parts that are most likely to contain behaviorally relevant information, allowing a more efficient allocation of the computationally draining resources of object (shape) recognition.

Another important property of vision is view-invariant object recognition. Traditional approaches to the problem have centered on an analysis of luminance edges and object shape. Overlooking the role of color might be a mistake. In the present report we have focused on a first-order analysis of the colors of objects, treating the pixels within an object as independent samples, selecting out one pixel per image as representative of all pixels, or averaging over them to obtain a single index for the color of an object. This analysis shows how important color could be for object vision, but color may provide even more information if one considers that the distribution of colors within an object might distinguish one object from another even if two objects have the same average color. This “color index” is not only relatively distinct but also largely immune to changes in the 2-D projection caused by differences in viewing geometry (Swain & Ballard, 1991). Object identification using color indexing can fail when the illumination changes, but these failures can be remedied by indexing RGB (or LMS) ratios instead of raw RGB signals (Funt & Finlayson, 1995). Computations of local cone ratios such as those implemented by double opponent cells in V1 (Conway, 2001) can support color constancy (Foster & Nascimento 1994), giving rise to color representations that are stable against changes in illumination. Taken together, these results suggest that the output of V1 made available to IT would enable IT to extract chromatic signatures from objects that could support view-invariant shape calculations. Thus color could

contribute not only to the calculation of valence but also recognition.

Keywords: objects, color, scene statistics, machine vision, neurophysiology

Acknowledgments

We thank the intramural research program of the NIH for financial support, June Kim and Shridhar Singh for help collecting and analyzing the data, and Nico Kriegeskorte for providing images used in Figure 6.

Commercial relationships: none.

Corresponding author: Bevil R. Conway.

E-mail: bevil@nih.gov.

Address: Laboratory of Sensorimotor Research, National Eye Institute, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA.

References

- Bartels, A., & Zeki, S. (2000). The architecture of the colour centre in the human visual brain: New results and a review. *The European Journal of Neuroscience*, *12*(1), 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. *Cerebral Cortex*, *9*(3), 257–263.
- Bohon, K. S., Hermann, K. L., Hansen, T., & Conway, B. R. (2016). Representation of perceptual color space in macaque posterior inferior temporal cortex (the V4 complex). *eNeuro*, *3*(4):e0039–16.2016, 1–28.
- Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*(1), 1–34.
- Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(27), 1–27, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Chiu, W. C., & Fritz, M. (2015). See the difference: Direct pre-image reconstruction and pose estimation by differentiating HOG. *2015 IEEE International Conference on Computer Vision, arXiv: 1505.00663*, 468–476.
- Conway, B. R. (2001). Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V-1). *Journal of Neuroscience*, *21*(8), 2768–2783.
- Conway, B. R. (2018). The organization and operation of inferior temporal cortex. *Annual Review of Vision Science*, *4*, 381–402, <https://doi.org/10.1146/annurev-vision-091517-034202>.
- Conway, B. R., Moeller, S., & Tsao, D. Y. (2007). Specialized color modules in macaque extrastriate cortex. *Neuron*, *56*(3), 560–573.
- Conway, B. R., & Tsao, D. Y. (2005). Color architecture in alert macaque cortex revealed by fMRI. *Cerebral Cortex*, *16*(11), 1604–1613.
- Conway, B. R., & Tsao, D. Y. (2009). Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex. *Proceedings of the National Academy of Science, USA*, *106*(42), 18034–18039.
- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol 1, Proceedings, 1*, 886–893.
- Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *Journal of Physiology*, *357*, 241–265.
- DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Science*, *11*(8), 333–341.
- Fairchild, M. D. (2001). A revision of CIECAM97s for practical applications. *Color Research and Application*, *26*(6), 418–427.
- Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2010). Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(9), 1627–1645.
- Finlay, V. (2014). *The brilliant history of color in art*. Los Angeles, CA: Getty Publications, J. Paul Getty Museum.
- Foster, D. H., & Nascimento, S. M. C. (1994). Relational colour constancy from invariant cone-excitation ratios. *Proceedings of the Royal Society of London, B: Biological Sciences*, *257* (115–121).
- Funt, B. V., & Finlayson, G. D. (1995). Color constant color indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*(5), 522–529.
- Gauthier, I., & Tarr, M. J. (2016). Visual object recognition: Do we (finally) know more now than we did? *Annual Review of Vision Science*, *2*, 377–396.
- Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*, *10*(13), 805–808.

- Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., . . . Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences, USA*, 114(40), 10785–10790.
- Harada, T., Goda, N., Ogawa, T., Ito, M., Toyoda, H., Sadato, N., Komatsu, H. (2009). Distribution of colour-selective activity in the monkey inferior temporal cortex revealed by functional magnetic resonance imaging. *The European Journal of Neuroscience*, 30(10), 1960–1970.
- Huang, Z. L., Pan, Z. X., & Lei, B. (2017). Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data. *Remote Sensing*, 9(9), 1–21, <https://doi.org/10.3390/rs9090907>.
- Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, 76(6), 1210–1224.
- Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences, USA*, 107(25), 11163–11170.
- Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, 97(6), 4296–4309.
- Koenderink, J. J. (2010). The prior statistics of object colors. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 27(2), 206–217.
- Komatsu, H. (1998). Mechanisms of central color vision. *Current Opinion in Neurobiology*, 8(4), 503–508.
- Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *The Journal of Neuroscience*, 33(25), 10235–10242.
- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2: 4.
- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., . . . Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6), 1126–1141.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90, <https://doi.org/10.1145/3065386>.
- Lafer-Sousa, R., & Conway, B. R. (2013). Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nature Neuroscience*, 16(12), 1870–1878.
- 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. *The Journal of Neuroscience*, 36(5), 1682–1697.
- 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. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 29(5), 657–670.
- Lee, S. U., & Ju, B. K. (2017). Wide-gamut plasmonic color filters using a complementary design method. *Science Reports*, 7: 40649.
- Liu, T., Sun, J., Zheng, N.-N., Tang, X., & Shum, H.-Y. (2007). Learning to detect a salient object. *2007 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–8), Minneapolis, MN, <https://doi.org/10.1109/CVPR.2007.383047>.
- MacLeod, D. I., & Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America*, 69(8), 1183–1186.
- Mammeri, A., Boukerche, A., Feng, J. W., & Wang, R. F. (2013). North-American speed limit sign detection and recognition for smart cars. *Proceedings of the 2013 38th Annual IEEE Conference on Local Computer Networks* (pp. 154–161), Sydney, NSW, <https://doi.org/10.1109/LCNW.2013.6758513>.
- Mely, D. A., Kim, J., McGill, M., Guo, Y., & Serre, T. (2016). A systematic comparison between visual cues for boundary detection. *Vision Research*, 120, 93–107.
- Moroney, N., & Beretta, G. (2011). Validating large-scale lexical color resources. *HP Laboratories Technical Report, HPL-2011-226 AIC Midterm Meeting, 2011*, Zurich.
- Naselaris, T., Stansbury, D. E., & Gallant, J. L. (2012). Cortical representation of animate and inanimate objects in complex natural scenes. *Journal of Physiology, Paris*, 106(5–6), 239–249.
- Paxinos, G., Huang, X-F., & Toga, A. W. (2000). *The rhesus monkey brain in stereotaxic coordinates*. San Diego, CA: Academic Press.
- Regan, B. C., Julliot, C., Simmen, B., Vienot, F., Charles-Dominique, P., & Mollon, J. D. (2001).

- Fruits, foliage and the evolution of primate colour vision. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 356(1407), 229–283.
- Rhodes, L., Argersinger, M. E., Gantert, L. T., Friscino, B. H., Hom, G., Pikounis, B., . . . Rhodes, W. L. (1997). Effects of administration of testosterone, dihydrotestosterone, oestrogen and fadrozole, an aromatase inhibitor, on sex skin colour in intact male rhesus macaques. *Journal of Reproduction and Fertility*, 111(1), 51–57.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11), 1019–1025.
- Rowland, H. M., & Burriss, R. P. (2017). Human colour in mate choice and competition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1724): 20160350.
- Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., Connolly, A. C. (2015). The animacy continuum in the human ventral vision pathway. *Journal of Cognitive Neuroscience*, 27(4), 665–678.
- Stansbury, D. E., Naselaris, T., & Gallant, J. L. (2013). Natural scene statistics account for the representation of scene categories in human visual cortex. *Neuron*, 79(5), 1025–1034.
- Stephen, I. D., Law Smith, M. J., Stirrat, M. R., & Perrett, D. I. (2009). Facial skin coloration affects perceived health of human faces. *International Journal of Primatology*, 30(6), 845–857.
- Stoughton, C. M., & Conway, B. R. (2008). Neural basis for unique hues. *Current Biology*, 18(16), R698–R699.
- Swain, M. J., & Ballard, D. H. (1991). Color Indexing. *International Journal of Computer Vision*, 7(1), 11–32.
- Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences*, 5(5), 211–215.
- Tanigawa, H., Lu, H. D., & Roe, A. W. (2010). Functional organization for color and orientation in macaque V4. *Nature Neuroscience*, 13(12), 1542–1548.
- Tootell, R. B., Nelissen, K., Vanduffel, W., & Orban, G. A. (2004). Search for color ‘center(s)’ in macaque visual cortex. *Cerebral Cortex*, 14(4), 353–363.
- Verhoef, B. E., Bohon, K. S., & Conway, B. R. (2015). Functional architecture for disparity in macaque inferior temporal cortex and its relationship to the architecture for faces, color, scenes, and visual field. *The Journal of Neuroscience*, 35(17), 6952–6968.
- Webster, M. A., Mizokami, Y., & Webster, S. M. (2007). Seasonal variations in the color statistics of natural images. *Network*, 18(3), 213–233.
- Webster, M. A., & Mollon, J. D. (1997). Adaptation and the color statistics of natural images. *Vision Research*, 37(23), 3283–3298.
- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, arXiv:1411.1792, 3320–3328.
- Zamir, S. W., Vazquez-Corral, J., & Bertalmio, M. (2017). Gamut extension for cinema. *IEEE Transactions on Image Processing*, 26(4), 1595–1606.
- Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (Eds.). *Computer Vision - ECCV 2014. ECCV 2014. Lecture Notes in Computer Science*, vol 8689. Switzerland - Cham: Springer, https://doi.org/10.1007/978-3-319-10590-1_53.
- Zhang, J., Barhomi, Y., & Serre, T. (2012). A New Biologically Inspired Color Image Descriptor. *Computer Vision - Eccv 2012, Pt V*, 7576, 312–324.