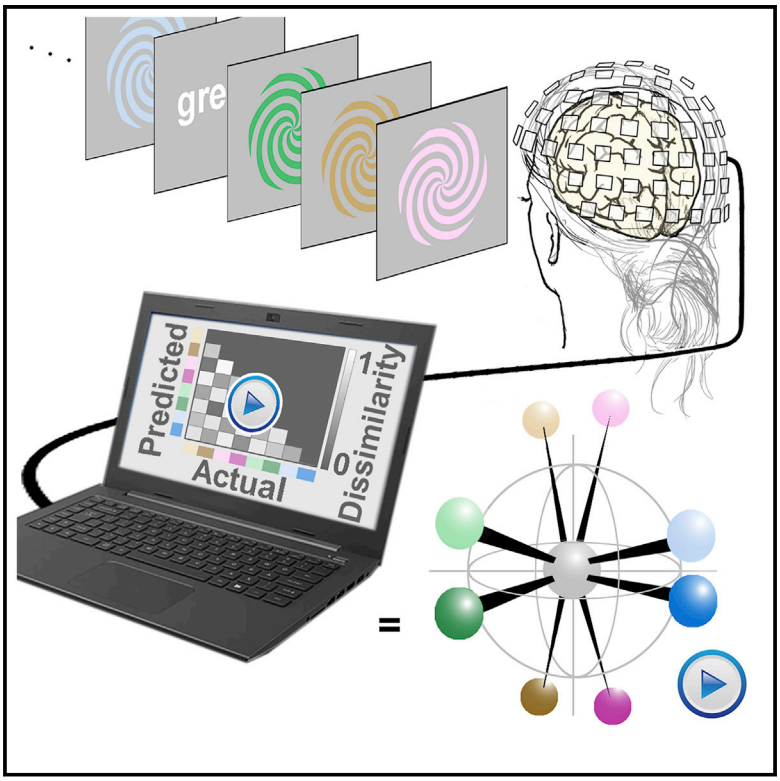
Article

Current Biology

Color Space Geometry Uncovered with Magnetoencephalography

Graphical Abstract



Authors

Isabelle A. Rosenthal,
Shridhar R. Singh,
Katherine L. Hermann,
Dimitrios Pantazis, Bevil R. Conway

Correspondence

bevil@nih.gov

In Brief

Rosenthal, Singh, Hermann, et al. use magnetoencephalography to decode color. By analyzing the similarity relationships among responses to colors varying in hue and luminance, the authors discover a dynamic geometry of the neural representation of color. The geometry predicts universal color-naming patterns and makes new predictions for perception.

Highlights

- Stimulus color can be decoded from surface magnetoencephalography (MEG) recordings
- Perceptual representations give rise to semantic representations, but not the reverse
- The results reveal a neural geometry of color space that is dynamic
- The geometry explains universal color-naming patterns and generates new hypotheses

Rosenthal et al., 2021, Current Biology *31*, 515–526
February 8, 2021 Published by Elsevier Inc.
https://doi.org/10.1016/j.cub.2020.10.062

Current Biology

CellPress

Article

Color Space Geometry Uncovered with Magnetoencephalography

Isabelle A. Rosenthal,[1,4,6] Shridhar R. Singh,[1,6] Katherine L. Hermann,[1,5,6] Dimitrios Pantazis,[2] and Bevil R. Conway[1,3,7,*]
[1]Laboratory of Sensorimotor Research, National Eye Institute, Building 49, NIH Main Campus, Bethesda, MD 20892, USA
[2]McGovern Institute for Brain Research, Massachusetts Institute of Technology, 524 Main Street, Cambridge, MA 02139, USA
[3]National Institute of Mental Health, Bethesda, MD 20892, USA
[4]Present address: Computation and Neural Systems Program, California Institute of Technology, Pasadena, CA, USA
[5]Present address: Department of Psychology, Stanford University, Stanford, CA, USA
[6]These authors contributed equally
[7]Lead Contact
*Correspondence: bevil@nih.gov
https://doi.org/10.1016/j.cub.2020.10.062

SUMMARY

The geometry that describes the relationship among colors, and the neural mechanisms that support color vision, are unsettled. Here, we use multivariate analyses of measurements of brain activity obtained with magnetoencephalography to reverse-engineer a geometry of the neural representation of color space. The analyses depend upon determining similarity relationships among the spatial patterns of neural responses to different colors and assessing how these relationships change in time. We evaluate the approach by relating the results to universal patterns in color naming. Two prominent patterns of color naming could be accounted for by the decoding results: the greater precision in naming warm colors compared to cool colors evident by an interaction of hue and lightness, and the preeminence among colors of reddish hues. Additional experiments showed that classifiers trained on responses to color words could decode color from data obtained using colored stimuli, but only at relatively long delays after stimulus onset. These results provide evidence that perceptual representations can give rise to semantic representations, but not the reverse. Taken together, the results uncover a dynamic geometry that provides neural correlates for color appearance and generates new hypotheses about the structure of color space.

INTRODUCTION

A long-standing goal of color science has been the development of a uniform color space in which the perceived difference between any two colors is captured by the Euclidian distance separating the colors.[1] Many strategies have been employed,[2] including Munsell,[3] CIELUV/CIELAB/CIECAM02,[4,5] the Natural Color System,[6] and the Optical Society of America Color System.[7] But there is no geometry that predicts perceptual relationships over both short and long intervals within the space,[1,4] and while the sequence of colors is consistent across different spaces, the geometrical spacing of colors differs. These differences reflect task demands[8] and choices among competing priorities, for example, accurately representing discrimination thresholds versus appearance.[9] The diversity of spaces suggests that there is no unique geometry of color and has exposed mysteries about how the underlying neural mechanisms work.[10,11]

Stimulus color can be decoded from functional magnetic resonance imaging activity[12] and from responses of color-tuned neurons in monkeys.[13,14] Moreover, the extent to which activity patterns correspond to the sequence of hues in color spaces can be used to identify neural populations and brain areas that are likely involved in encoding color.[12,14] The present work continues this tradition, using tools of neuroscience to understand how color perception is implemented. But we also explore whether the traditional logic can be flipped: rather than using neuroscience to understand how known aspects of perception arise, can direct neural measurements help resolve questions about perception such as the geometrical spacing of colors? Our goal is to reverse-engineer a geometry that captures the relationships among colors from direct measurements of brain responses and to determine how color representations in the brain unfold over time.

We take up our work using MEG, which affords fine temporal precision, coupled with a decoding approach.[15,16] We first show that stimulus color can be decoded using multivariate analyses of the spatial pattern of responses across the array of MEG sensors, confirming preliminary findings by ourselves and others.[17–23] We then examine the similarity relationships among the patterns of neural responses elicited by different colors. We judged colors to be more similar if they elicited more similar patterns of MEG activity. This approach is analogous to the way perceptual similarities among colors can be established, where participants might be asked to pick a color that is most similar to a target. We use stimuli with equal absolute cone contrast along the two cone-opponent dimensions that the retina uses to encode color (Figure 1A).[24,25] Because MEG is thought to predominantly reflect cortical responses,

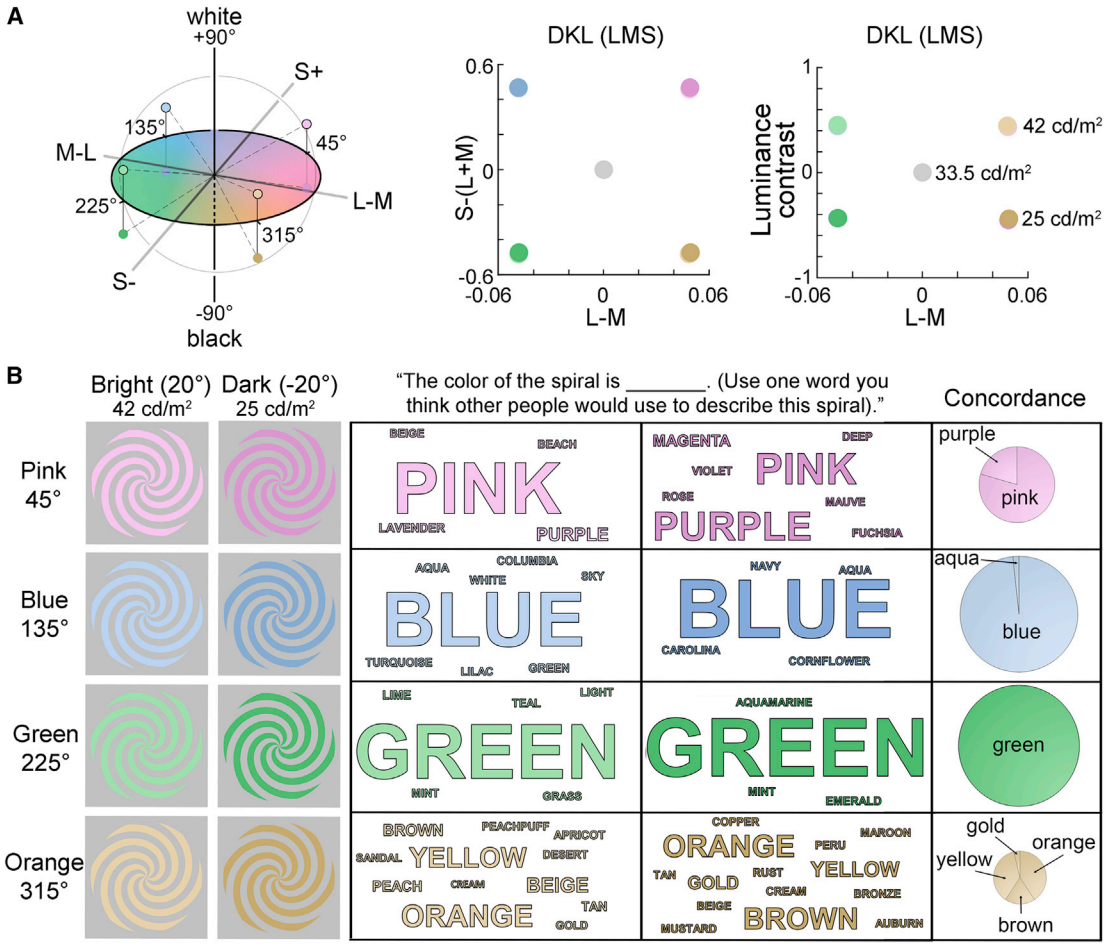


Figure 1. Specification of the Stimulus Colors and How They Are Named

(A) Left: diagram of the color space used to specifiy the colors. The space is defined by the cardinal cone-opponent mechanisms (L versus M, and S; where L, M, and S are activations of the three cone types).[24,25] The stimulus colors (symbols) were defined by the intermediate axes. The right panels show the cone and luminance contrast of the stimuli, relative to the gray adapting background.

(B) Left: configuration of the stimuli; spirals subtended 10° of visual angle and varied in hue (rows) and luminance contrast (columns). Center: word clouds, reflecting the distribution of color names assigned to the stimuli. Font size shows how many of the 70 subjects assigned that word to the stimulus. The smallest words (e.g., "lavender") represent just one response; the largest words ("green" and "blue" in dark hues column) represent 66 responses. Data were obtained from 52 volunteers who performed the experiments online, and the 18 participants from whom we collected MEG data. Right: concordance of color names used across luminance contrast. The size of the pie chart reflects the proportion of participants who used the same word across luminance sets, and the content of chart indicates which words were used for both luminance sets. Almost all participants used "green" for both high and low luminance contrast versions of that hue (large pie for hue 225°), whereas relatively few participants used the same term for both high and low luminance contrast versions of the yellow/brown/ochre stimulus (small pie for color 315°), and among these participants, several terms were used (many pie slices for 315°). The concordance across luminance sets, computed as the percentage of people that used the same term for both light and dark versions of the hue was pink = 55.6%; blue = 85.7%; green = 88.6%; orange = 38.6%. The most commonly used term for hue 135° (blue) at the light|dark luminance level was used by 87%|94% of participants; the most commonly used term for hue 225° (green) was used by 90%|94% of participants. For hue 45° (pink), there was lower consensus (81%|50%) and lower concordance (44% used pink for light and dark versions; 11% used purple). For hue 315° (orange), there was no clear consensus: the most commonly used term, orange, was only used by 30% of participants (at both luminance levels). There was no difference in the level of concordance between blue and green (chi-square test of proportions, p = 0.61; Bonferroni alpha = 0.0083); similarly, there was no difference in concordance between pink and orange (p = 0.04; Bonferroni alpha = 0.0083).

the experiments therefore reveal the transformation of retinal signals by the cortex.

We evaluate the approach by relating the patterns of neural activity elicited by different colors to universal patterns in color naming, guided by the hypothesis that the cross-linguistic patterns in color naming derive from an underlying geometry of the neural representation. We consider two universal aspects of color naming. First, there is a greater precision in naming warm colors compared to cool colors.[26–28] This color-naming efficiency reflects an interaction of hue and luminance: among the basic color categories (red, pink, orange, yellow, brown, green, blue, and purple), the cool colors (green and blue) are not distinguished from each other by differences in lightness, whereas the warm colors are,[29] notably yellow and brown.[30] Second, the hue associated with red is the most salient chromatic color, reflected in the preeminence of red among basic

color terms across languages.[29,31] Finally, we measured MEG responses to color words to compare perceptual and semantic representations.

RESULTS

We measured MEG responses to eight colored spirals. The stimuli comprised four hues (two warm colors and two cool colors) at two luminance-contrast levels (light and dark, defined by higher and lower luminance relative to the adapting background). The absolute cone contrast was equated for all the stimuli: stimuli were modulated along the intermediate axes in cone-opponent color space (Figure 1A). This space, referred to as DKL or MB-DKL after its inventors,[24,25] has two cardinal chromatic axes: the L-M axis, along which colors vary in activation of L and M cones and maintain constant S-cone activity; and the S axis, along which colors vary in activation of S cones and maintain constant L-cone and M-cone activity.[32] We also measured responses to achromatic text of the words "green" and "blue." The experimental paradigm represents a compromise between having a large enough representation of color space, a set of colors with well-defined and balanced cone activation, and a small enough set of stimuli to enable each stimulus to be presented a large number of times to ensure sufficient power for decoding analysis. Except for the words, the stimuli had the same shape and uniform texture, ensuring the same pattern of retinotopic activation. Thus the extent to which patterns of activation elicited by one spiral versus another can be decoded must be attributed to differences in color.

The experiment was designed to test for four possible asymmetries in the geometry of the neural representation of color: along the L-M cardinal axis (warm versus cool, see Figure S2, and Lindsey and Brown[29]), the S axis (S increments versus S decrements), between the intermediate axes (orange-blue versus green-pink), and across lightness levels (light versus dark). To evaluate the approach, we compared the pattern of neural responses to color-naming behavior. As expected, participants often used different terms for the light and dark versions of warm colors and the same term for light and dark versions of the cool colors (Figure 1B). The naming data of the participants in the MEG experiments were supplemented with data collected online (see STAR Methods for a comparison of the two datasets). The percentage of participants who used the same term for both versions of the given hue (concordance across luminance sets) was 55.6% (for pink), 85.7% (for blue), 88.6% (for green), and 38.6% (for orange).

Classifying Colors Using MEG Responses

MEG activity elicited by the colored spirals was measured in 18 volunteers. Each spiral was presented 500 times—the large number of trials was intended to ensure enough power for the decoding analyses (Figure S1A, inset, shows data reliability). For each of the 18 participants, at each time point relative to stimulus onset, a classifier was trained to decode stimulus color from the pattern of activity across the array of MEG sensors (using separate data to train and test the classifiers). Averaged across participants, color decoding reached significance at 55 ms (Figure 2A; 95% confidence interval, CI: [20, 70]); peak decoding accuracy was at 115 ms (95% CI: [105, 125]). The decoding time course showed a prominent initial peak (arrowhead, Figure 2A) corresponding to decoding stimulus onset, and a second hump (235 ms; [190, 265]) that we attribute to decoding stimulus cessation. Across subjects, the time to peak decoding ranged between 95 and 260 ms. This range reflects the fact that, while most participants showed maximal decoding corresponding to stimulus onset, three people showed maximal decoding corresponding to stimulus cessation. The performance of the classifiers cannot be attributed to confounds introduced by eye position or pupil diameter (Figure 2A; decoding by pupil diameter was insignificant; decoding by eye position was significant at long delays after stimulus onset, but with low decoding accuracy and only after decoding by MEG returned close to baseline; in all graphs of decoding, time points of significant decoding are indicated by the line of dots above the x axis). Pink was decodable with the shortest latency, regardless of luminance level (Figure 2B; lines of significance dots for each color are stacked top to bottom by onset latency). The relatively short latency for decoding pink is also evident when the data were analyzed relative to the time of peak decoding determined separately for each participant, which takes account of individual differences in decoding time (Figure S1B).

The magnitude of classification performance does not correspond to the absolute effect size but is a valid metric for relative effect sizes within a study.[33] At peak, classification accuracy varied by hue (repeated-measures two-way ANOVA on rank transformed data, $p = 0.005$), but not by luminance level ($p = 0.3697$; Figure 2C). Moreover, the magnitude of classification accuracy showed a subtle interaction of hue and luminance contrast, such that for warm colors, the light versions were more accurately classified compared to the dark versions, while for cool colors, the dark versions were more accurately classified compared to the light versions (the dashed contours are shifted to the left of the solid contours; Figure 2C; repeated-measures two-way ANOVA, no main effects, significant interaction between warm-cool and luminance contrast, $p = 0.0055$). This interaction was only evident for the 15-ms window spanning peak decoding (Video S1).

Post hoc tests for the four possible asymmetries in the geometry of the neural representation, assessed by differences in magnitude of color decoding across four axes through DKL color space, showed no evidence for an asymmetry in classifying colors along the L-M axis ($p = 0.28$; Bayes Factor B10 = 0.35) or across lightness levels ($p = 0.42$, B10 = 0.39). The results provided only anecdotal support for the null hypotheses, of no asymmetry in L-M and no asymmetry across lightness levels (the B10 values are not below 0.3; by convention, <0.3 indicates moderate evidence in favor of the null hypothesis; while >3 indicates moderate evidence against the null hypothesis[34]). The results provided moderate evidence for an asymmetry along the S axis ($p = 0.014$; B10 = 5.4; classification of S-increment colors was higher compared to S-decrement colors) and moderate evidence for an asymmetry between the pairs of intermediate colors ($p = 0.006$; B10 = 6.4; classification was worse for orange and blue compared to pink and green). But across the temporal evolution of the response, the most notable difference in the magnitude of decoding accuracy among the colors was for pink, which showed higher classification accuracy compared to luminance-matched hues (Figure 2d; Video S1).

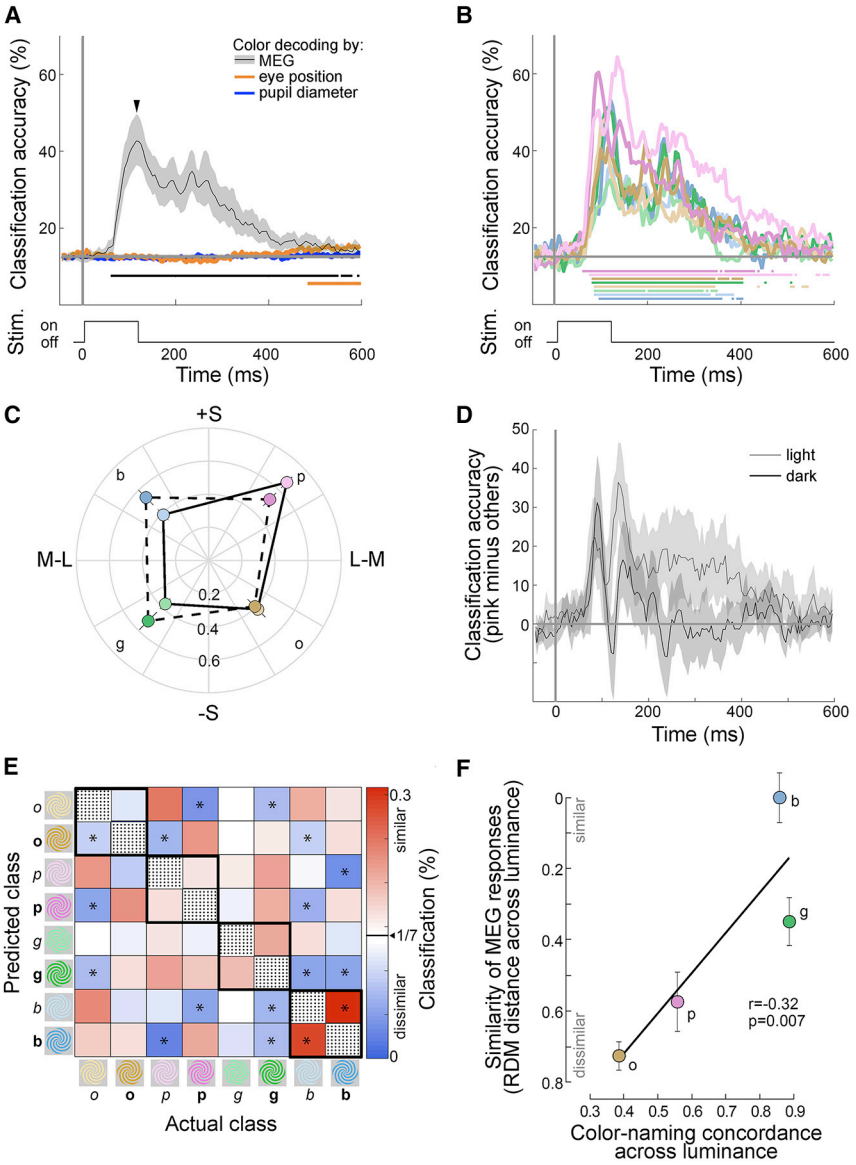


Figure 2. Decoding Stimulus Color from MEG Data

(A) Average classification accuracies across time for decoding the color of the eight spirals, averaged over all eight classification problems (N = 18 participants; chance = 1/8; shading shows 95% CI computed across participants). Step at the bottom shows the stimulus duration (0 ms = stimulus onset); arrowhead indicates the time point of peak decoding across all problems and participants. The sequence of data points forming a line above the x axis shows the times at which the classifier guessed the stimulus color above chance for at least five consecutive 5-ms bins (bootstrapped 1,000 times; significance of individual time bins was false-discovery-rate (FDR) corrected, $p < 0.05$).

(B) Average classification accuracy for decoding each of the eight colors (line color corresponds to stimulus color); other conventions as for panel (A). Peak decoding accuracy and onset latency of decoding varied among colors (see text for statistics); Figure S1 shows curves relative to peak decoding time rather than relative to stimulus onset.

(C) Peak decoding accuracy of each color, plotted as the radius length (b, blue; p, pink; o, orange; g, green). Decoding accuracies were determined from each subject at their overall individual peak decoding time. Error bars = SEM; Video S1 shows data for all time points relative to the time of peak decoding.

(D) Difference in classification accuracy amplitude for pink compared to the three other hues: $[(D_p - D_o) + (D_p - D_g) + (D_p - D_b)]/3$, where D_p, D_o, D_g, and D_b are the magnitudes of decoding accuracy for pink, orange, green, and blue, computed separately for light and dark colors; shading shows 95% CI determined by bootstrapping (n = 1,000) across participants.

(E) Representational dissimilarity matrix (RDM) showing the classification forced error at each subject's individual peak decoding time, averaged across subjects. The colors in the heatmap indicate how often the classifier would assign a given spiral identity (vertical axis) when tested with data obtained by presenting a given spiral (horizontal axis), as a percentage such that the seven values in each column add to 1. The classifier was trained in the same manner as in (A), but forced to guess the most likely spiral besides the correct option. White on the color bar = chance level (1/7). Asterisks indicate that the forced error was either significantly above chance (redder) or below chance (bluer; $p < 0.05$; bootstrapped 95% CIs; n = 1,000).

(F) Correlation of the dissimilarity in neural representation across luminance contrast versus color-naming concordance across luminance contrast. The results show that light and dark versions of a hue are more likely to elicit less similar neural representations when they are labeled with a different color name (e.g ,yellow/brown). The data from (E) were converted to a triangular matrix by averaging the upper and lower triangular vectors; the values in the resulting triangular matrix were normalized to 1 and then subtracted from 1 so that 0 means maximally "similar" and 1 means maximally "dissimilar"; error bars = SEM. See also Figure S1 and Video S1.

Representational Similarity

The spatial patterns of neural activity elicited by different stimuli could show many different similarity relationships, even if the magnitude of decoding accuracy were the same for all the stimuli.[35] To determine the information encoded by the spatial pattern of responses, we used representational similarity analysis. Our goal was to determine the extent to which similarity relationships among patterns of neural responses elicited by colors reflect properties of perceptual color space, providing another access point to the geometry of the neural representation of color. Relationships in color appearance that underlie color space can be measured by having participants identify among a set of colors the color that is most similar, but not identical, to a given sample. We applied this logic to analyze the MEG data by asking what color the classifier picks when forced to pick any color except the correct color. In other words, what color besides the test color elicits a spatial pattern of neural responses that is most like the one caused by the test color? The

pattern of errors (Figure 2E) provides an estimate of the similarity in the neural representation among colors. Colors that show more similar neural response patterns correspond to a closer relationship in the neural geometry of color space (we also performed an analysis on the errors of the classifier performance, which yielded similar results, but the number of errors is relatively low compared to the number of trials; the forced-error analysis has more power and is more directly analogous to similarity matching paradigms used in psychophysics, although psychophysical experiments would typically use many more colors).

The heatmap shown in Figure 2E is largely symmetric about the identity diagonal, confirming that the pattern of results is meaningful (Pearson correlation coefficient comparing the two halves of the matrix, $r = 0.9665$, $p = 7\times10^{-17}$). For example, the confusion between colors of the same hue but different luminance were almost identical: dark orange was predicted when light orange was the actual color with almost the same likelihood that light orange was predicted when dark orange was the actual color, and so on (e.g., see each of the four black-outlined squares along the inverse diagonal). The cross-diagonal symmetry shows that the magnitude of the classification errors is informative, even when some errors are by themselves not significant. The confusion between stimuli of the same hue but different luminance varied among the four hues. For example, the pattern of MEG activity for blue was similar for light and dark versions of blue, while the pattern of MEG activity for orange was dissimilar for light and dark versions of orange. Rank-ordered by hue, the similarity relationships across luminance level corresponded to the concordance rates for color naming across luminance level (Figure 2F; correlation coefficient = -0.32, $p = 0.007$; the distance between hues across luminance contrast was greater for warm colors compared to cool colors, Kruskal-Wallis test, $p = 5\times10^{-4}$). The results show that the neural representation of warm colors is more impacted by luminance contrast than the neural representation of cool colors, which is consistent with the color-naming patterns shown in Figure 1B, in which blue and green at different luminance levels were typically assigned the same labels while pink and orange at different luminance levels were often given different labels.

Video S2 shows how the MEG dissimilarity matrix (DSM) of the forced-error analysis changes over time. The DSM was generated by averaging the symmetric halves of the error matrix. Figure 3 quantifies across time the correspondence of the DSM to three models, two of which reflect properties of color space recovered in perceptual similarity judgements: the clear separation of hues by luminance level (top panel, Figure 3A);[36] and the asymmetry in the impact of luminance contrast on the representation of warm versus cool colors (top panel, Figure 3B). The third model derives from observations of the data that, to our knowledge, have not been tested in perceptual experiments: an asymmetry in the representation of warm versus cool colors at the same luminance level (top panel, Figure 3C).

Representations for hues of a given luminance contrast were more similar than representations for hues across luminance levels (light and dark hues were separable), and this pattern existed for a continuous period of over 300 ms (Figure 3A, bottom panel). Representations of the light and dark versions of each of the cool colors were more similar compared to the warm colors, but over a shorter duration compared to the representation of luminance contrast (Figure 3B, bottom panel). And representations of the two warm colors were more similar than representations of the two cool colors within a luminance level (Figure 3C, bottom panel). That the three models were significantly correlated with the DSM over different durations shows that the geometry of the neural representation is dynamic.

Multidimensional scaling (MDS) provides a picture of the similarity relationships. The MDS plots are not simply a replotting of the dissimilarity matrices, but instead show the best spatial representation of the data as constrained by a limited number of dimensions. In the MDS plots, data points that are closer together correspond to neural representations that are more similar; i.e., pairs of colors that are closer together are more likely to be mistaken for each other if the classifier fails. Each MDS plot shows the similarity relationships among colors for a single snapshot in time relative to peak decoding time in each subject. Figure 3D shows the MDS plots constrained by two dimensions (MDS-2D) at six time points, three spanning times around the peak decoding of stimulus onset (-5 ms, 0 ms, and 5 ms) and three corresponding to times around the peak decoding of stimulus cessation (105 ms, 110 ms, and 115 ms). The observations derived from the MDS panels in Figure 3D are also evident in MDS plots constrained by three dimensions (Videos S3 and S4; we show the MDS-2D panels because they have the advantage that they are easier to reproduce on the page).

The MDS plots show three striking patterns. First, each MDS plot can be oriented such that the light and dark hues are displaced vertically, and in this configuration the sequence of hues aligns across luminance levels. Second, the MDS plots corresponding to peak decoding of stimulus cessation are distinguished by a dramatic separation of the hues by luminance contrast. Third, the similarity of colors, as assessed by the similarity of each light color to light pink, is consistent with the sequence in psychological color spaces, pink-orange-green-blue (Figure 3D, inset, shows pink at each time point [top row], and at each time point, the color that was most similar to pink [second row], and the color that was second-most similar to pink [third row], and finally, the color that was least similar to pink [bottom row]). The MDS plots also reflect the observation that the neural representation for light and dark warm colors was less similar than the representation of light and dark cool colors (e.g., the two pink dots are further apart than the two blue dots), and the observation that, within a given luminance level, the two warm colors showed more similar representations than the two cool colors (e.g., the light pink and light orange dots are closer together than the light green and light blue dots). Again, the dynamic structure of the MDS plots supports the conclusion that the geometry of the neural representation is not static.

Distinct Representations of Colors and Color Words

The use of color-naming data to evaluate patterns of neural responses elicited by colors prompted a question: is the geometry of the neural representation of color determined, or influenced, by implicit linguistic or category associations? To address this question, we randomly interleaved trials with the words "green" and "blue," written in white text, and evaluated whether classifiers trained on the pattern of MEG activity elicited

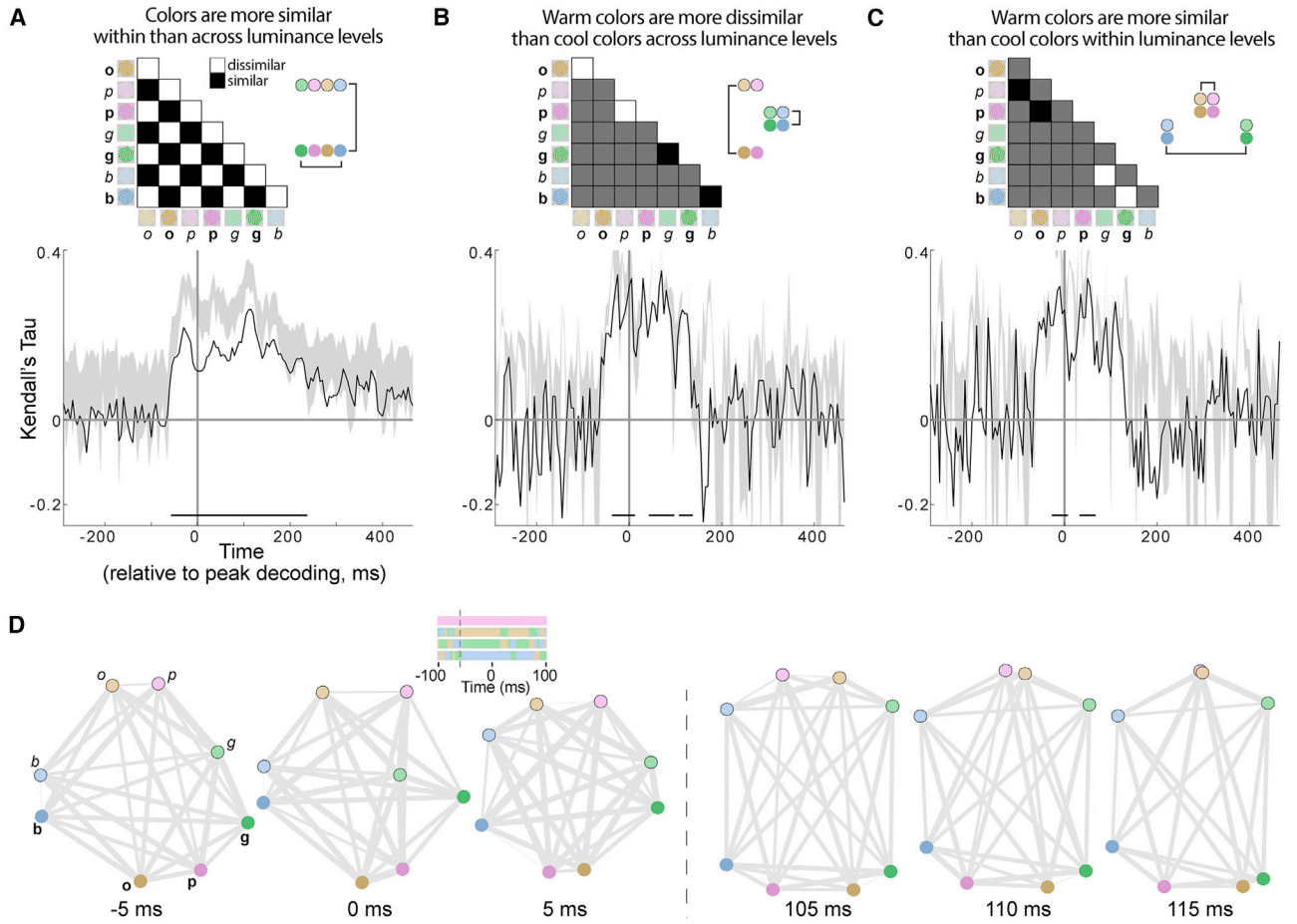


Figure 3. Similarity Relationships among Colors Determined by MEG

(A) Quantification of the observation that patterns of MEG activity for hues of a given luminance level were, on average, more similar than patterns of activity for hues across luminance levels. The triangle shows the model against which the representational dissimilarity matrix (RDMs) was compared at each point in time. The RDMs across time were generated from the confusion matrix of the forced error classifier's performance. The lower and upper triangular vector of the matrices (see Figure 2E) were averaged and scaled from 0 to 1. These values were subtracted from 1 to produce an RDM that quantifies the dissimilarity of color pairs. The graph shows the correlation (Kendall's Tau) of the model and the RDM. Gray shading shows the upper and lower bound of the noise ceiling—the highest correlation of the true model RDM prediction. The horizontal line of data points above the x axis shows time points where the correlation was significant for at least five consecutive 5-ms bins (significance of individual time bins was FDR corrected). Video S1 shows the RDM across time. The graphic to the right of the model provides a visual representation of the model and can be compared to the multidimensional scaling (MDS) representations in (D).

(B) Quantification of observation that patterns of MEG activity elicited by warm colors across luminance levels were more dissimilar than those patterns elicited by cool colors across luminance levels. In computing the model correlations, only values highlighted in black and white were considered. Other conventions as in (A).

(C) Quantification of the patterns of MEG activity for warm colors within one luminance level were more similar than patterns of MEG activity for cool colors within one luminance level. Other conventions as in (B).

(D) MDS plots of the RDM data, at six time points relative to the time point of peak decoding accuracy: 0 ms corresponds to peak decoding of stimulus onset, 110 ms corresponds to peak decoding of stimulus cessation, and the vertical dashed line separates MDS plots corresponding to onset from those corresponding to stimulus cessation. Symbol color corresponds to the color of the stimulus, and the light versions are outlined in black to further distinguish the data points. Shorter distances correspond to higher levels of classification confusion (higher similarity). The area of each line is proportional to the dissimilarity in the representation of the colors joined by the line. Hues are separated by luminance contrast (light colors are at the top in each panel) and arranged by hue in a sequence, from left to right, of blue, orange (or pink), pink (or orange), and green. Fit of MDS for each panel, left to right: Pearson's; $R = 0.84$, $p = 2x10^{-8}$; $R = 0.84$, $p = 3x10^{-8}$; $R = 0.85$, $p = 2x10^{-8}$; $R = 0.91$, $p = 1x10^{-11}$; $R = 0.94$, $p = 9x10^{-14}$; $R = 0.95$, $p = 7x10^{-15}$. The inset shows, top to bottom, the sequence of light colors in order of similarity to pink, determined by the three-dimensional MDS, at each point in time; 0 ms is peak decoding; the vertical dashed line shows when decoding became significant. For most of the significant decoding time course, orange is most similar to pink, followed by green, and blue is least similar to pink. See also Videos S2, S3, and S4 for three-dimensional MDS.

by the words could decode color from MEG activity elicited by green and blue spirals. We chose "green" and "blue" because color-naming concordance across luminance levels was highest for green and blue spirals. We reasoned that the pattern of MEG activity elicited by these words would therefore have the greatest likelihood of predicting patterns of activity elicited by colored spirals. We assessed cross-temporal generalization of the MEG patterns to test three hypotheses. First, that the representation instantiated by color is identical to the one instantiated by color terms. If true, we expect significant decoding performance

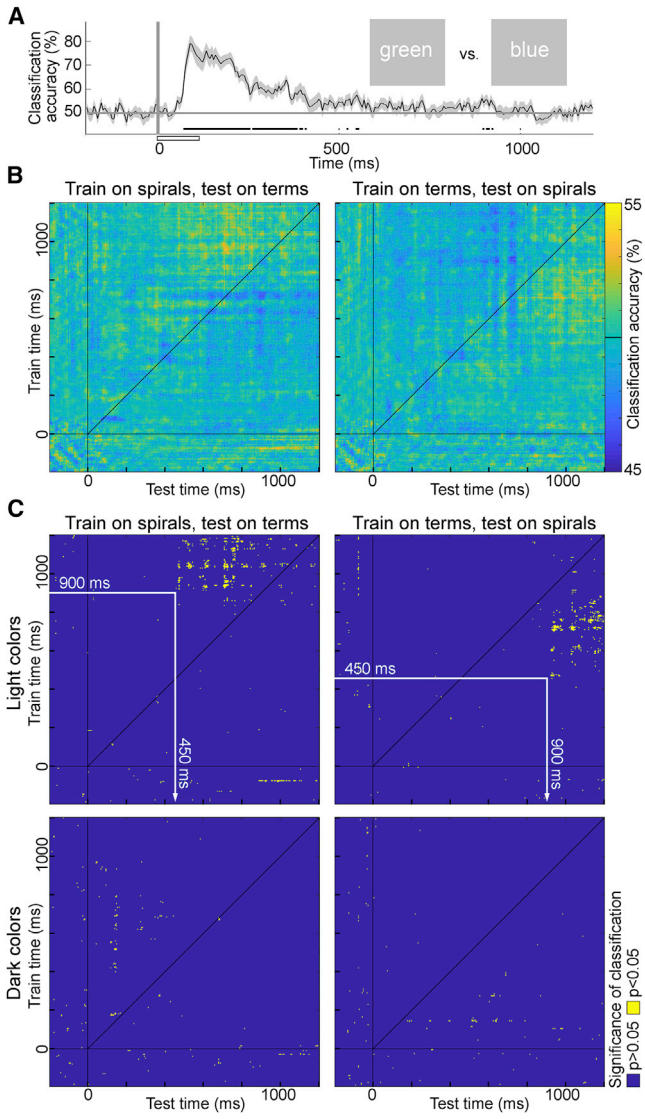


Figure 4. Decoding Color Appearance from Representations of Color Words

(A) Average accuracy of classifiers (N = 18 participants) decoding whether the term "green" or "blue" was presented. Plotting conventions as in Figure 2A; open box above the zero shows the stimulus duration.

(B) Cross-temporal, cross-stimulus decoding. Left: Average accuracy of classifiers trained on color spirals and tested on color terms (accuracies averaged across both luminance levels of spirals). At every 5-ms bin of the training data, a classifier was trained and accuracy (color on the heatmap) was measured at every 5-ms bin of test data. Right: classifiers trained on color term data and tested on color spiral data; same conventions as left. The color terms were presented as white text on a gray background; the color spirals were light and dark versions of green and blue spirals.

(C) Significance tests of the cross-temporal decoding plots for classifiers trained (and tested) on the light versions of the colored spirals (top panels) and the dark versions of the colored spirals (bottom panels), FDR corrected. Significant activation (yellow) is evident in the top panels; the left and right panels show symmetric significant time points across the diagonal (white arrow; e.g., the onset of significant decoding in the top left occurs at training times ~900 ms and testing times of ~450 ms).

along the diagonal in plots where we train (or test) on representations elicited by colors and test (or train) on representations elicited by terms. Second, that perceptual representations of color give rise to semantic representations. If true, we expect significance above the diagonal in plots where we train on representations elicited by colors (y axis) and test on representations elicited by terms (x axis). Third, that semantic representations can drive perceptual representations. Participants were asked to visualize the color corresponding to each term; thus, if the hypothesis is true, we expect significance above the diagonal in plots where we train on representations elicited by terms (y axis) and test on representations elicited by colors (x axis).

Unsurprisingly, term identity (green or blue) could be readily decoded from the pattern of MEG activity elicited by presentation of the terms (onset of 65 ms [95% CI: 65, 70]; peak accuracy of 80% [74, 86] obtained at 100 ms [90, 120]; Figure 4A). This is unsurprising because visual cortex is expected to respond to differences in the retinotopic projection of the two words. Classifiers trained on MEG data elicited by colored spirals could not decode term identity from MEG activity elicited by words at any time delay between training and testing over the period of time during which stimulus color was readily decoded (up to about 400 ms after stimulus onset). Similarly, during this time period, classifiers trained on MEG data obtained using the words could not decode the color of the spirals. These results do not support the first hypothesis and suggest that the representation instantiated by words and colors are distinct. At much longer time delays, classifiers trained on the neural representation elicited by colors could decode neural representations of words that emerged relatively earlier (training times after ~900 ms, testing times after ~450 ms; Figure 4B), supporting the second hypothesis. But classifiers trained on the neural representation elicited by words could not decode neural representations of colors emerging relatively earlier, which does not support the third hypothesis. The asymmetry about the diagonal in the cross-temporal decoding plots suggests that the representation of a color gives rise to a representation similar to one elicited by a color term, but that the representation of a color term does not subsequently elicit a representation similar to one generated by color perception. Notably, the cross-stimulus decoding was only significant when classifiers were trained (or tested) using data elicited by the light-colored spirals (Figure 4C). The results of the cross-stimulus (color/color-word) experiments suggest that the geometry of the neural representation of color discovered by analyzing responses to colored spirals is not confounded by implicit linguistic or category associations.

Temporal Dynamics of the Representation of Color

Decoding stimulus color was significant for a substantial amount of time, spanning about 300 ms (Figure 2B). To examine the temporal dynamics of the neural representation, we evaluated the extent to which classifier decoding generalized across time.[19,15,37] Can classifiers trained at one point in time decode activity at another point in time? If the representation consists of a dynamic chain, with neural activity progressing through a series of unique states, classification performance will only be possible when test and train data are from the same time point. Alternatively, if representations are relatively stable, it should be

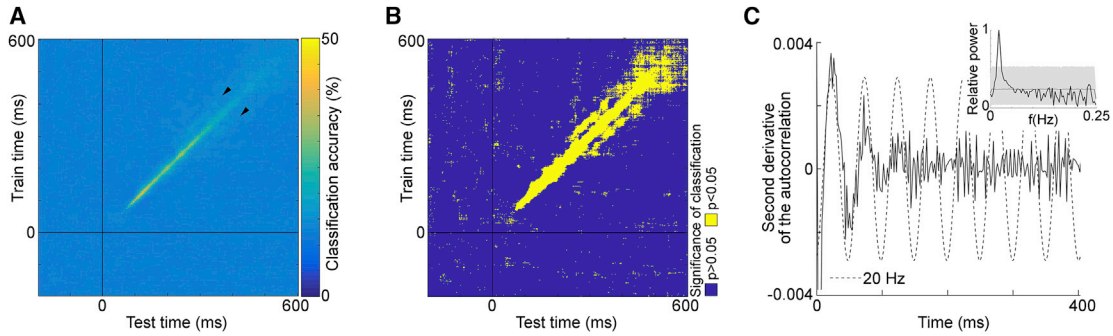


Figure 5. Cross-Temporal Generalization of Color Identity

(A) Mean classification accuracy across participants (N = 18) in classifying the eight color spirals. In addition to the strong classification accuracy along the diagonal, the plot shows faint bands of significant classification accuracy parallel to the diagonal (arrowheads).

(B) Plot showing time points of significant decoding, estimated by bootstrapping ($p < 0.05$: FDR corrected).

(C) Analysis of the spatial pattern in A) to determine the temporal profile of the parallel bands. We performed a two-dimensional autocorrelation analysis on cross-temporal decoding accuracies by shifting the cross-temporal decoding map horizontally along the test-time axis. We computed the two-dimensional Pearson's correlation coefficient of the original map with each of the shifted maps and performed a Fourier analysis on the second derivatives of the two-dimensional correlation coefficients. The plot shows the second derivatives with a sine wave that has a frequency equal to the peak of the single-sided amplitude spectrum. The inset shows the single-sided amplitude spectrum of the second derivative time series; shading shows the 95% CI of the scrambled second derivatives plot.

possible to classify activity obtained at one time point using training data obtained at a different time point.

Figure 5A shows the cross-temporal generalization analysis for spiral identity, and Figure 5B indicates time points where decoding performance was above chance. Decoding performance was strongest along the diagonal, consistent with the first possibility. But the plot also shows evidence of faint bands of significant decoding performance parallel to the diagonal, which is consistent with some version of the second possibility, specifically one in which a pattern of neural activity is reactivated at some later time point. To determine the time course of this reactivation, we reproduced the cross-temporal generalization analysis using 2-ms bins and performed a two-dimensional autocorrelation analysis on the cross-temporal generalization plot. For the analysis, we restricted the train times to a window of 275–600 ms. We shifted the temporal generalization plot horizontally along the test-time axis and computed the two-dimensional Pearson's correlation coefficient between the original plot and each of the shifted plots. We computed the second derivatives of the correlation coefficients and performed a Fourier analysis on the resulting time series. The frequency with the maximum amplitude was 20 Hz, which corresponds to a reactivation time of 50 ms (Figure 5C).

DISCUSSION

This paper shows that stimulus color can be decoded from MEG data and presents an approach for investigating the similarity relationships among colors using multivariate analyses of MEG responses. The success of the decoding analysis requires that different colors elicit different spatial patterns of response across the brain. The results therefore provide evidence not only of a topographic representation of color in the human brain, as predicted by neurophysiological recordings in monkeys,[38–41] but also that MEG has the spatial resolution to detect the topography. The spatial patterns of neural responses elicited by different colors were compared using representational similarity analysis. We interpret the similarity relationships as evidence of a geometry of the neural representation of color, which, we speculate, derives from the color statistics of objects[42] and the environment[43] and accounts for perceptual similarity judgements that drive many universal patterns in color naming. Our next step is to measure MEG responses to more colors, at many saturation and lightness levels, which will enable the construction of a complete color solid in which the spacing of the colors corresponds to equal distances in the dissimilarity among neural responses.

The geometry recovered by the neural data is consistent with general features of perceptual color space as reflected in MDS analysis of perceptual similarity judgements:[36] representations of luminance contrast were clearly separable (Figure 3A), and the sequence of hues was as predicted from perception (Figure 3d, inset). In addition, the results uncovered two prominent properties of the cortical representation of color related to color-naming patterns. First, the neural representation elicited by stimuli of the same hue across luminance levels were more dissimilar for warm colors compared to cool colors (Figure 3B). This pattern of results correlated with the extent to which participants used the same term for both light and dark versions of each hue (Figure 2F). For example, participants were more likely to use different terms for the light and dark versions of pink than they were to use different terms for the light and dark versions of green; and the patterns of MEG activity elicited in response to light and dark pink were more dissimilar than the patterns of MEG activity elicited in response to light and dark green. The MEG results show that patterns of activity elicited by warm colors were more precisely encoded and therefore more accurately decoded compared to cool colors, which may account for the greater naming efficiency of warm colors.[26,28] Second, regardless of luminance contrast, pink was decoded with the highest decoding accuracy over much of the temporal evolution of the response (Figure 2D) and with the shortest onset latency

(Figure 2B; Figure S1B). This result cannot be attributed to differences in contrast of the stimuli (the stimuli were all matched in absolute cone contrast). The result echoes observations in monkeys showing that reddish hues produce very high gamma oscillations[44] and provides a neural correlate for the preeminence of reddish hues in color naming.[29,31]

The present work aimed to determine neural correlates of some known properties of color perception. But we also sought to explore whether the neural data might provide new hypotheses about perception. For example, we found that neural representations for colors at a given luminance level were more similar between the two warm colors than between the two cool colors. To our knowledge, color-naming studies have not shown (or explicitly tested) whether colors at a given luminance level are less efficiently communicated for warm colors compared to cool colors. We wonder whether the result reflects the adaptation of visual processing to the color statistics of objects: warm colors, regardless of the specific hue, are associated with parts of scenes we label as objects,[26] and color tuning across inferior temporal cortex (IT) is correlated with object-color probability.[42]

The MEG data were obtained using colored stimuli defined by the cone-opponent mechanisms thought to be implemented by bipolar cells of the retina. Because MEG reflects predominantly cortical activity, any asymmetries in the MEG responses provide clues to the transformation by the cortex of the subcortical representation of color. An analysis of the magnitude of color decoding uncovered two asymmetries in the neural representation of color space. First, at the time point of peak decoding, classification accuracy was slightly lower for colors corresponding to the daylight locus (orange-blue) compared to the anti-daylight locus (pink-green; Figure 2C). This observation may reflect differences in response magnitude in V1 to different mixtures of subcortical channels[45,46] and may have implications for understanding neural mechanisms of color constancy.[46–50] For example, achromatic settings show the greatest variation for colors along the daylight locus,[51] and discrimination of illumination changes is worst for spectral changes aligned with the daylight locus.[48] And second, decoding accuracy was higher for S-increment colors compared to S-decrement colors (Figure 2C); this reflects an asymmetry evident in neurophysiological recordings of the retina,[52,53] V1,[54] and extrastriate cortex.[55,56]

During times in which stimulus color could be readily decoded (up to ~400 ms after stimulus onset), stimulus color could not be decoded from patterns of MEG activity elicited by color terms presented in achromatic (white) text. But classifiers trained on representations elicited by colored spirals at relatively long delays (~900 ms after stimulus onset) could decode representations elicited by words that emerged earlier (~450 ms after presentation of terms). Curiously, the cross-stimulus decoding was only significant using classifiers trained (or tested) using data elicited by the light versions of the colored spirals (Figure 4C). We speculate that the use of white text may have biased participants to visualize the light version of the color term. Classifiers trained using representations elicited by words could not decode representations elicited by colors at relatively earlier delays. This is unlike classifiers trained on responses elicited by colors, which can decode color implied by color-diagnostic shapes.[19] Thus it seems that color-diagnostic shapes, but not color words, can elicit neural representations similar to those elicited by real colors. The results presented here are consistent with the idea that the brain processes information in a directional way,[57] with perceptual representations capable of driving semantic or cognitive representations, but not the reverse—semantic representations seem incapable of instantiating perceptual representations.[58]

The plots of cross-temporal generalization of spiral color (Figure 5) show a pattern of bands parallel to the identity diagonal, which provides a rare example (the first, to our knowledge) of a theoretical possibility—the reactivation pattern of King and Dehaene.[15] This pattern suggests that the perceptual representation of color is encoded by a unique sequence of brain states that is reactivated with a slight delay. The reactivation time was 50 ms, implying that the brain encodes color information with recurrent activation loops of 20 Hz. We speculate that these reactivation loops involve the hierarchical network of color-biased regions within the ventral visual pathway[59–62] that connect primary visual cortex eventually with frontal cortex.[63,64] The plots of cross-temporal generalization of spiral color may also provide evidence of the emergence of relatively late semantic representation: note the distinctive square of significant decoding at the upper right corner of Figure 5B, which starts at about 450 ms, about the same time that cross-stimulus decoding became significant in Figure 4C.

In sum, the present results provide clues to the neural basis that governs the geometric relationships among colors and to the connection between perceptual representations and color naming. Although no physiological measurement can decode qualia, the results nonetheless establish a proof of concept that multivariate analysis of MEG activity can recover similarity relationships in how colors appear. The results underscore the existence of neural representations of color in which hue cannot be completely decoupled from luminance contrast. Moreover, the results imply that the cortex transforms the retinal representation of color to generate a perceptual representation characterized by multiple asymmetries across color space, which unfold in a dynamic way following stimulus onset. That color depends on complex and dynamic neural representations is consistent with the existence of an extensive network of cortical areas implicated in processing color[62,65] and may explain why it has been so difficult to establish a uniform color space. In the words of Deane Judd,[1] the pursuit of such a space may be misguided, even if it continues to have intuitive and practical value.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Visual Stimuli
  - Controlling for luminance artifacts
  - Color-naming

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.10.062.

ACKNOWLEDGMENTS

We thank Chris Baker and members of his laboratory, Susan Wardle, Alex Rehding, and Daniel Garside for helpful discussions, and Christine Vonder Haar for help collecting and analyzing the MTurk data. This research was supported in part by the Intramural Research Program of the National Institutes of Health, National Eye Institute, and grant R01 EY-0233223 to B.R.C.

AUTHOR CONTRIBUTIONS

K.L.H. and B.R.C. conceived and designed the experiments; I.A.R. and K.L.H. collected the data; I.A.R., S.R.S., K.L.H., and B.R.C. analyzed the data; I.A.R., S.R.S., and B.R.C. made the figures; D.P. provided expertise and resources for MEG; B.R.C. supervised the work and wrote the paper; all authors provided comments on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 13, 2020
Revised: September 21, 2020
Accepted: October 21, 2020
Published: November 16, 2020

REFERENCES

1. Judd, D.B. (1968). Ideal Color Space: The super-importance of hue differences and its bearing on the geometry of color space. Palette *30*, 21–28.
2. Kuehni, R., and Schwartz, A. (2008). Color Ordered: A Survey of Color Systems from Antiquity to the Present (New York, NY: Oxford University press).
3. Munsell, A.H. (1907). A Color Notation, a measured color system, based on the three qualities Hue, Value, and Chroma, Second Edition (Boston, Geo.: H. Ellis Co.).
4. Cheung, V. (2012). Uniform color spaces. In Handbook of Visual Display Technology, J. Chen, W. Cranton, and M. Fihn, eds. (Berlin, Heidelberg: Springer).
5. Fairchild, M.D. (2013). Color Appearance Models, 3rd (Wiley-IS&T).
6. Shamey, R., Shepherd, S., Abed, M., Chargualaf, M., Garner, N., Dippel, N., Weisner, N., and Kuehni, R.G. (2011). How well are color components of samples of the Natural Color System estimated? J. Opt. Soc. Am. A Opt. Image Sci. Vis. *28*, 1962–1969.
7. Boynton, R.M., and Olson, C.X. (1990). Salience of chromatic basic color terms confirmed by three measures. Vision Res. *30*, 1311–1317.
8. Bae, G.Y., Olkkonen, M., Allred, S.R., and Flombaum, J.I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. J. Exp. Psychol. Gen. *144*, 744–763.
9. Brainard, D.H., and Stockman, A. (2010). Colorimetry. In OSA Handbook of Optics, 3rd edition, M. Bass, ed. (McGraw-Hill), pp. 10.11–11.56.
10. Smet, K.A., Webster, M.A., and Whitehead, L.A. (2016). A simple principled approach for modeling and understanding uniform color metrics. J. Opt. Soc. Am. A Opt. Image Sci. Vis. *33*, A319–A331.
11. Ennis, R.J., and Zaidi, Q. (2019). Geometrical structure of perceptual color space: Mental representations and adaptation invariance. J. Vis. *19*, 1.
12. Brouwer, G.J., and Heeger, D.J. (2009). Decoding and reconstructing color from responses in human visual cortex. J. Neurosci. *29*, 13992–14003.
13. Zaidi, Q., Marshall, J., Thoen, H., and Conway, B.R. (2014). Evolution of neural computations: Mantis shrimp and human color decoding. Iperception *5*, 492–496.
14. Bohon, K.S., Hermann, K.L., Hansen, T., and Conway, B.R. (2016). Representation of Perceptual Color Space in Macaque Posterior Inferior Temporal Cortex (the V4 Complex). eNeuro *3*, https://doi.org/10.1523/ENEURO.0039-16.2016.
15. King, J.R., and Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn. Sci. *18*, 203–210.
16. Grootswagers, T., Wardle, S.G., and Carlson, T.A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. J. Cogn. Neurosci. *29*, 677–697.
17. Hermann, K., Pantazis, D., and Conway, B.R. (2015). The dynamics of color processing in humans measured with MEG. Society for Neuroscience Annual Meeting, 790.03/M32.
18. Rosenthal, I.A., Hermann, K.L., Vonder Haar, C., Pantazis, D., and Conway, B.R. (2017). Decoding hue and luminance with magnetoencephalography. Society for Neuroscience annual meeting, 774.03.
19. Teichmann, L., Grootswagers, T., Carlson, T.A., and Rich, A.N. (2019). Seeing versus knowing: The temporal dynamics of real and implied colour processing in the human brain. Neuroimage *200*, 373–381.
20. Sandhaeger, F., von Nicolai, C., Miller, E.K., and Siegel, M. (2019). Monkey EEG links neuronal color and motion information across species and scales. eLife *8*, https://doi.org/10.7554/eLife.45645.
21. Hermann, K., Rosenthal, I., Singh, S., Pantazis, D., and Conway, B.R. (2020). Temporal dynamics of the neural mechanisms for encoding hue and luminance contrast uncovered by magnetoencephalography. bioRxiv. https://doi.org/10.1101/2020.06.17.155713.
22. Rosenthal, I., Singh, S., Hermann, K., Pantazis, D., and Conway, B.R. (2020). Uncovering the geometry of color space with magnetoencephalography (MEG). bioRxiv. https://doi.org/10.1101/2020.08.10.245324.
23. Hajonides, J.E., Nobre, A.C., van Ede, F., and Stokes, M.G. (2020). Decoding visual colour from scalp electroencephalography measurements. bioRxiv. https://doi.org/10.1101/2020.07.30.228437.
24. MacLeod, D.I., and Boynton, R.M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. J. Opt. Soc. Am. *69*, 1183–1186.
25. Derrington, A.M., Krauskopf, J., and Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. J. Physiol. *357*, 241–265.
26. Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S.T., and Conway, B.R. (2017). Color naming across languages reflects color use. Proc. Natl. Acad. Sci. USA *114*, 10785–10790.
27. Zaslavsky, N., Kemp, C., Regier, T., and Tishby, N. (2018). Efficient compression in color naming and its evolution. Proc. Natl. Acad. Sci. USA *115*, 7937–7942.
28. Conway, B.R., Ratnasingam, S., Jara-Ettinger, J., Futrell, R., and Gibson, E. (2020). Communication efficiency of color naming across languages provides a new framework for the evolution of color terms. Cognition *195*, 104086.
29. Lindsey, D.T., and Brown, A.M. (2006). Universality of color names. Proc. Natl. Acad. Sci. USA *103*, 16608–16613.
30. Buck, S.L. (2015). Brown. Curr. Biol. *25*, R536–R537.

31. Berlin, B., and Kay, P. (1969). Basic color terms: their universality and evolution (Berkeley, CA: University of California Press).
32. Zaidi, Q., and Halevy, D. (1993). Visual mechanisms that signal the direction of color changes. Vision Res. *33*, 1037–1051.
33. Hebart, M.N., and Baker, C.I. (2018). Deconstructing multivariate decoding for the study of brain function. Neuroimage *180* (Pt A), 4–18.
34. Keysers, C., Gazzola, V., and Wagenmakers, E.J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. Nat. Neurosci. *23*, 788–799.
35. Wardle, S.G., Taubert, J., Teichmann, L., and Baker, C.I. (2020). Rapid and dynamic processing of face pareidolia in the human brain. Nat. Commun. *11*, 4518.
36. Indow, T. (1988). Multidimensional studies of Munsell color solid. Psychol. Rev. *95*, 456–470.
37. Hebart, M.N., Bankson, B.B., Harel, A., Baker, C.I., and Cichy, R.M. (2018). The representational dynamics of task and object processing in humans. eLife *7*, https://doi.org/10.7554/eLife.32816.
38. Conway, B.R., and 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. USA *106*, 18034–18039.
39. Garg, A.K., Li, P., Rashid, M.S., and Callaway, E.M. (2019). Color and orientation are jointly coded and spatially organized in primate primary visual cortex. Science *364*, 1275–1279.
40. Liu, Y., Li, M., Zhang, X., Lu, Y., Gong, H., Yin, J., Chen, Z., Qian, L., Yang, Y., Andolina, I.M., et al. (2020). Hierarchical Representation for Chromatic Processing across Macaque V1, V2, and V4. Neuron. https://doi.org/10.1016/j.neuron.2020.07.037.
41. Roe, A.W., Chelazzi, L., Connor, C.E., Conway, B.R., Fujita, I., Gallant, J., et al. (2012). Towards a unified theory of visual area V4. Neuron *74*, 12–29.
42. Rosenthal, I., Ratnasingam, S., Haile, T., Eastman, S., Fuller-Deets, J., and Conway, B.R. (2018). Color statistics of objects, and color tuning of object cortex in macaque monkey. J. Vis. *18*, 1.
43. McDermott, K.C., and Webster, M.A. (2012). Uniform color spaces and natural image statistics. J. Opt. Soc. Am. A Opt. Image Sci. Vis. *29*, A182–A187.
44. Shirhatti, V., and Ray, S. (2018). Long-wavelength (reddish) hues induce unusually large gamma oscillations in the primate primary visual cortex. Proc. Natl. Acad. Sci. USA *115*, 4489–4494.
45. Goddard, E., Mannion, D.J., McDonald, J.S., Solomon, S.G., and Clifford, C.W. (2010). Combination of subcortical color channels in human visual cortex. J. Vis. *10*, 25.
46. Lafer-Sousa, R., Liu, Y.O., Lafer-Sousa, L., Wiest, M.C., and 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 Opt. Image Sci. Vis. *29*, 657–670.
47. Delahunt, P.B., and Brainard, D.H. (2004). Does human color constancy incorporate the statistical regularity of natural daylight? J. Vis. *4*, 57–81.
48. Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G.D., and Hurlbert, A. (2014). Chromatic illumination discrimination ability reveals that human colour constancy is optimised for blue daylight illuminations. PLoS ONE *9*, e87989.
49. Winkler, A.D., Spillmann, L., Werner, J.S., and Webster, M.A. (2015). Asymmetries in blue-yellow color perception and in the color of 'the dress'. Curr. Biol. *25*, R547–R548.
50. Lafer-Sousa, R., Hermann, K.L., and Conway, B.R. (2015). Striking individual differences in color perception uncovered by 'the dress' photograph. Curr. Biol. *25*, R545–R546.
51. Chauhan, T., Perales, E., Xiao, K., Hird, E., Karatzas, D., and Wuerger, S. (2014). The achromatic locus: effect of navigation direction in color space. J. Vis. *14*, https://doi.org/10.1167/14.1.25.
52. Dacey, D.M., and Lee, B.B. (1994). The 'blue-on' opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. Nature *367*, 731–735.
53. Tailby, C., Solomon, S.G., and Lennie, P. (2008). Functional asymmetries in visual pathways carrying S-cone signals in macaque. J. Neurosci. *28*, 4078–4087.
54. Conway, B.R., and Livingstone, M.S. (2006). Spatial and temporal properties of cone signals in alert macaque primary visual cortex. J. Neurosci. *26*, 10826–10846.
55. Conway, B.R. (2014). Color signals through dorsal and ventral visual pathways. Vis. Neurosci. *31*, 197–209.
56. Wandell, B.A., Poirson, A.B., Newsome, W.T., Baseler, H.A., Boynton, G.M., Huk, A., Gandhi, S., and Sharpe, L.T. (1999). Color signals in human motion-selective cortex. Neuron *24*, 901–909.
57. Siuda-Krzywicka, K., Witzel, C., Bartolomeo, P., and Cohen, L. (2020). Color Naming and Categorization Depend on Distinct Functional Brain Networks. Cereb. Cortex. https://doi.org/10.1101/2020.04.13.038836.
58. Siuda-Krzywicka, K., Witzel, C., Taga, M., Delanoe, M., Cohen, L., and Bartolomeo, P. (2019). When colours split from objects: The disconnection of colour perception from colour language and colour knowledge. Cogn. Neuropsychol. https://doi.org/10.1080/02643294.2019.1642861.
59. Beauchamp, M.S., Haxby, J.V., Jennings, J.E., and 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.
60. Lafer-Sousa, R., and Conway, B.R. (2013). Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. Nat. Neurosci. *16*, 1870–1878.
61. Lafer-Sousa, R., Conway, B.R., and 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.
62. Conway, B.R. (2018). The Organization and Operation of Inferior Temporal Cortex. Annu. Rev. Vis. Sci. *4*, 381–402.
63. Bird, C.M., Berens, S.C., Horner, A.J., and Franklin, A. (2014). Categorical encoding of color in the brain. Proc. Natl. Acad. Sci. USA *111*, 4590–4595.
64. Haile, T.M., Bohon, K.S., Romero, M.C., and Conway, B.R. (2019). Visual stimulus-driven functional organization of macaque prefrontal cortex. Neuroimage *188*, 427–444.
65. Siuda-Krzywicka, K., and Bartolomeo, P. (2020). What Cognitive Neurology Teaches Us about Our Experience of Color. Neuroscientist *26*, 252–265.
66. Meyers, E.M. (2013). The neural decoding toolbox. Front. Neuroinform. *7*, 8.
67. Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., and Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. PLoS Comput. Biol. *10*, e1003553.
68. Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., and Leahy, R.M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. Comput. Intell. Neurosci. *2011*, 879716.
69. Kleiner, M., Brainard, D.H., and Pelli, D. (2007). What's new in Psychtoolbox-3? Perception *36*, 1–16.
70. Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. *10*, 437–442.
71. Mannion, D.J., McDonald, J.S., and Clifford, C.W. (2009). Discrimination of the local orientation structure of spiral Glass patterns early in human visual cortex. Neuroimage *46*, 511–515.
72. Seymour, K., Clifford, C.W., Logothetis, N.K., and Bartels, A. (2010). Coding and binding of color and form in visual cortex. Cereb. Cortex *20*, 1946–1954.
73. Brouwer, G.J., and Heeger, D.J. (2013). Categorical clustering of the neural representation of color. J. Neurosci. *33*, 15454–15465.
74. Westland, S., Ripamonti, C., and Cheung, V. (2012). Computational colour science using matlab, 2nd Edition (John Wiley and Sons).

75. Brainard, D.H. (1996). Cone contrast and opponent modulation color spaces in *Human Color Vision*, Second Edition, P.K. Kaiser, and R.M. Boynton, eds. (Washington, D.C.: Optical Society of America).

76. Hansen, T., and Gegenfurtner, K.R. (2006). Color scaling of discs and natural objects at different luminance levels. Vis. Neurosci. *23*, 603–610.

77. Stockman, A., and Sharpe, L.T. (2000). Tritanopic color matches and the middle- and long-wavelength-sensitive cone spectral sensitivities. Vision Res. *40*, 1739–1750.

78. Bradley, A., Zhang, X., and Thibos, L. (1992). Failures of isoluminance caused by ocular chromatic aberrations. Appl. Opt. *31*, 3657–3667.

79. Taulu, S., and Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys. Med. Biol. *51*, 1759–1768.

80. Isik, L., Meyers, E.M., Leibo, J.Z., and Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. J. Neurophysiol. *111*, 91–102.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Raw and analyzed data	This paper	https://neicommons.nei.nih.gov/#/
Software and Algorithms		
MEG color task paradigm and materials	This paper	N/A
Code for data analysis and decoding	This paper	N/A
MATLAB	Mathworks, Natick, MA, USA	https://www.mathworks.com/products/MATLAB.html; RRID: SCR_001622
Neural Decoding Toolbox	[66]	RRID:SCR_009012
Toolbox for Representational Similarity Analysis	[67]	RRID:SCR_019029
Brainstorm	[68]	http://neuroimage.usc.edu/brainstorm/; RRID:SCR_001761
Psychtoolbox	[69,70]	http://psychtoolbox.org/; RRID:SCR_002881SCR_001761

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and data should be directed to and will be fulfilled by the Lead Contact, BRC (bevil@nih.gov).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The datasets used in this study are available on openneuro https://doi.org/10.18112/openneuro.ds003352.v1.0.0. Datasets and code used to generate the analyses and figures will be available on NEICOMMONS, https://neicommons.nei.nih.gov

EXPERIMENTAL MODEL AND SUBJECT DETAILS

This study examined neurotypical human participants (N = 18, age 19-37 years, 11 female) using MEG. These participants had normal or corrected-to-normal vision, were right-handed, and spoke English as a first language. One participant was an author and thus not naive to the purpose of the study. During MEG participants' first session, they were screened for colorblindness using Ishihara plates.

In addition to the 18 participants analyzed in the main thrust of this study, two pilot versions of the experiment were deployed. The first examined 2 participants (age 20-30 years, 1 female) to determine the behavioral task and decoding parameters. The second pilot experiment measured responses in four participants (age 20-30; 3 female), 52 online participants (29 female, age 19-63 years), distinct from the participants tested in the laboratory, were recruited and tested using Amazon Mechanical Turk (mTurk) and provided monetary compensation for their participation. In addition to color answers, some basic demographic information was collected (age, gender, language spoken, nationality, education level, vision, colorblindness, normal sleep/wake times, and handedness). Participants were excluded if English was not their first language or if they reported color blindness.

All experimental procedures involving participants tested in laboratory were approved by the Wellesley College Institutional Review Boards, the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects, and the National Institutes of Health Intramural Institute Clinical Research Review Committee.

METHOD DETAILS

Visual Stimuli

Stimuli were eight square-wave spiral gratings on a neutral gray background (Figure 1A)[71–73] or achromatic words (white text on a gray background) of the words "green" and "blue" spanning 10 degrees of visual angle, centered on the display in front of the participant. The color words "green" and "blue" were chosen for the experiment because they were the terms used with the highest consensus in the color-naming experiment; we reasoned that if neural patterns elicited by color words and by colors were similar,

they would be most likely recovered by using the color words for stimuli that had the highest color naming consensus. Participants were asked to imagine the corresponding spiral color indicated by the term each time they saw the words. The 8 stimulus colors, four hues at two luminance-contrast levels, of matched cone contrast, were defined in DKL color space[24,25] using implementations by Westland[74] and Brainard:[75] the axes of this color space are defined in terms of activation of the two cone-opponent post-receptoral chromatic mechanisms (Figure 1A). The scaling of the x and y axes in the DKL space were constrained by the gamut of the monitor, as in other studies.[32,76] The display was gamma corrected and calibrated using a spectroradiometer (PR655, Spectrascan, Chatsworth CA). The z axis is defined by luminance. The four hues were defined by the intermediate axes of DKL space. For ease of communication, throughout the text we refer to the colors of the stimuli using color terms: at 45° (pink), 135° (blue), 225° (green), and 315° (orange). Our use of color terms should not be interpreted as implying that the stimuli are best exemplars of these terms, or that these terms are the consensus terms for communicating these colors. Two spirals – one high luminance (20° elevation; "light") and one low luminance (340° elevation; "dark")–were created at each hue. The neutral adapting background was 33.5 cd/m2. The luminance contrast of the stimuli was 26%. Modulation of the cone-opponent mechanisms, shown in Figure 1A, was computed relative to the adapting background gray, using the Stockman and Sharpe 2degree cone fundamentals,[77] Judd corrected.

Controlling for luminance artifacts

The experiments were designed to minimize the impact of luminance artifacts. Because the stimuli were presented on a luminance pedestal, either an increment or a decrement, any luminance artifact introduced either by chromatic aberration or by deviations of an individual's isoluminant setting would be diluted by the luminance contrast of the luminance pedestal. The colored spirals consisted of relatively large bands of color, corresponding to about 0.8 cycles per degree. The fraction of the display that would be subject to chromatic aberration—the amount of the display consisting of hard color-gray edges—is relatively small. The stimuli assumed acommon luminosity function for the participants. Differences in luminosity functions across people are unlikely to cause luminance contrasts that are greater in magnitude than the luminance contrast of the luminance pedestal. The variability in luminance settings could mean that colors at a given luminance-contrast level did not all have precisely the same luminance contrast, but because of the luminance pedestal, they should all have the same sign of luminance contrast. Individual differences causing subtle variations in luminance contrast among colors of a given luminance level might influence classifier accuracy for a given color in an induvial participant, but it is difficult to imagine how these individual differences could drive systematic patterns of results observed across the set of participants. The impact of luminance artifacts caused by chromatic aberration are especially pronounced at detection threshold.[78] All the stimuli we used were substantially above detection threshold. Finally, the stimuli were configured in the same spiral shape as the stimuli used in studies of fMRI decoding of color.[73]

The stimuli in the fMRI color-decoding study used stimuli that were isoluminant (i.e., not on a luminance pedestal) and would therefore be more likely than the stimuli used in the present experiments to be impacted by luminance artifacts. If such artifacts had a substantial impact on cortical responses, it should be possible to decode color from brain areas that are sensitive to luminance contrast but not tuned to color, because variability in color would be associated with systematic variation in luminance contrast. Brouwer and Heeger found that activity in none of the areas with known sensitivity to luminance contrast, including L01, L02, V3A/B or MT, could decode color. The method we used and that of Brouwer and Heeger (fMRI) differ (MEG versus fMRI), but both methods depend on differences in the spatial representation across the brain. The spatial sensitivity of fMRI is purportedly better than that of MEG, so we would expect the difference in methodology to favor fMRI for decoding based on luminance-contrast artifacts.

Color-naming

Color names for the eight spirals used in the MEG experiments were obtained from 52 participants tested online, and the 18 participants from the MEG experiment who were tested in person. We collected data using both in-lab and online populations to test whether the in-lab results would generalize across a more diverse population, and to increase power in the behavioral data.

The 18 MEG participants completed the color-naming task during the first of two MEG sessions using the same calibrated display as was used in the MEG experiments. Participants were shown each of the color stimuli (Figure 1) three times in random order and given the prompt "The color of the spiral is ____. (Use one word that you think other people would use to describe this spiral)." Participants spoke their responses to the researcher sitting outside the room via 2-way intercom. Participants controlled the duration that they saw each spiral via button press. Only one response per spiral per participant was used in the analysis (selected at random from among the three responses per stimulus provided by each participant). The spiral was presented on a neutral gray background, as large as the screen would allow. Before beginning the experiment, participants were presented with a consent screen notifying them that participation was anonymous and that the task could be completed only once. We recognize that using an online testing platform means that the stimuli would be displayed on different screens, with presumably different calibration settings, for different participants. The color naming results from these experiments would therefore be expected to have more variance than experiments obtained using the same calibrated display for all participants.

The results from the two sets of participants (online versus in-lab) were comparable and consistent with predictions based on the literature. For (MTurk|in-lab) participants, the same term was used at both luminance levels for hue 315° by (31%|61%); for hue 45° by (56%|56%); for hue 135° by (83%|94%); and for hue 225° by (85%|100%). The percentage of people who used the same term for the light and dark versions was the same for the MTurk and in-lab participants (Chi-square test of proportions, Bonferroni corrected for multiple comparisons alpha = 0.0125; for hue 45°, $p = 0.99$; for hue 135°, $p = 0.22$; for hue 225°, $p = 0.08$; for hue 315°, $p = 0.02$). These

results showed that participants in both experiments tended to use the same label for the high and low luminance version of each hue more often when labeling cool colors (hues 135° and 225°) than warm colors (hues 45° and 315°). The most commonly used terms for hues 135° and 225° were the same for the MTurk participants and the in-lab participants (blue, green; consistent across luminance contrasts). The most commonly used terms for hue 45° were pink|pink (light|dark MTurk participants) and pink|purple (in-lab); for hue 315°, yellow|yellow (MTurk); orange|brown (in-lab). Because the results were comparable between the two sets of participants, we combined the data in the analyses that relate the MEG results to color naming. The correlation computed in Figure 2F, using color-naming data for only the 18 people who participated in the MEG experiments, is $R = -0.29$, $p = 0.013$.

Throughout the report, we refer to the blue and green colors as "cool," and the pink and orange colors as "warm." This warm-cool designation corresponds to differences in cone modulation along the L-M axis. Although it is arguably the case that the warmest and coolest colors do not align with the poles of the L-M axis, the boundaries between warm and cool colors derived from color-naming data[29] fall very close to colors of the S-cone axis in DKL color space, consistent with the notion that "warm" and "cool" are defined by the extent of L-M modulation (Figure S2). Of course, dichromats distinguish warm and cool, so the L-M designation is not sufficient for warm-cool designation.

MEG Acquisition and Preprocessing

Participants were scanned in the Athinoula A. Martinos Imaging Center of the McGovern Institute for Brain Research at the Massachusetts Institute of Technology (MIT) over the course of 2 sessions, on an Elekta Triux system (306-channel probe unit consisting of 102 sensor triplets, with 204 planar gradiometer sensors, and 102 magnetometer sensors). Stimuli were back-projected onto a 44" screen using a SXGA+ 10000 Panasonic DLP Projector, Model No. PT-D10000U (50/60Hz, 120V) using custom code written in MATLAB using Psychtoolbox.[69,70] Data were recorded at a sampling rate of 1000Hz, filtered between 0.03-330Hz. Head location was recorded by means of 5 head position indicator (HPI) coils placed across the forehead and behind the ears. Before the MEG experiment began, 3 anatomical landmarks (bilateral preauricular points and the nasion) were registered with respect to the HPI coils, using a 3D digitizer (Fastrak, Polhemus, Colchester, Vermont, USA). During recording, pupil diameter and eye position data were collected simultaneously using an Eyelink 1000 Plus eye tracker (SR Research, Ontario, Canada) with fiber optic camera.

Once collected, raw data were preprocessed to offset head movements and reduce noise by means of spatiotemporal filters,[79] with Maxfilter software (Elekta, Stockholm). Default parameters were used: harmonic expansion origin in head frame = [0 0 40] mm; expansion limit for internal multipole base = 8; expansion limit for external multipole base = 3; bad channels omitted from harmonic expansions = 7 SD above average; temporal correlation limit = 0.98; buffer length = 10 s). In this process, a spatial filter was applied to separate the signal data from noise sources occurring outside the helmet, then a temporal filter was applied to exclude any signal data highly correlated with noise data over time. Following this, Brainstorm software[68] was used to extract the peri-stimulus MEG data for each trial (−200 to 600 ms around stimulus onset) and to remove the baseline mean.

MEG Task

After completing the in-lab color-naming task, MEG participants completed a 100-trial practice session of the 1-back task that would be used in the MEG experimental sessions. Once this was complete, participants were asked if they had any questions about the task or the experiment; eye-tracking calibration was performed; and MEG data collection began.

During stimulus presentation, participants were instructed to fixate at the center of the screen. Spirals were presented subtending 10° of visual angle, for 116 ms, centered on the fixation point, which was a white circle that appeared during inter-trial intervals (ITIs, 1 s). In addition to the spirals, the words "green" and "blue" were presented in white on the screen for the same duration, and probe trials to evaluate task performance were presented with a white ""?. Following each probe trial, which occurred every 3-5 stimulus trials (pseudorandomly interspersed, 24 per run), participants were instructed to report via button press if the two preceding spirals did or did not match in color (1-back hue task). Maximum response time was 1.8 s, but the trials advanced as soon as participants answered.

Participants were encouraged to blink only during probe trials, as blinking generates large electrical artifacts picked up by the MEG. Each run comprised 100 stimulus presentations, and participants completed 25 runs per session over the course of approximately 1.5 h. Between each run, participants were given a break to rest their eyes and speak with the researcher if necessary. Once 10 s had elapsed, participants chose freely when to end their break by button-press. Over the course of both sessions, participants viewed each stimulus 500 times. Individual runs were identical across subjects, but the order of runs was randomized between subjects. The sequence of stimuli within each run was random with the constraint that the total number of presentations was the same for each stimulus condition over the set of runs obtained for each participant.

In addition to the 18 participants analyzed in the main thrust of this study, two pilot versions of the experiment were deployed. The first, more limited pilot experiment was used to determine the behavioral task and decoding parameters. The data from these participants was used to choose the parameters for the decoding analysis used in the rest of the study (see below). The second pilot experiment was more extensive, involving four participants, four colors, and 500 stimulus presentations per stimulus. This experiment was used to evaluate the power for color decoding; the results of the second pilot experiment provide the first evidence, to our knowledge, of color decoding from MEG data, presented at the Society for Neurosciences Annual meeting in 2015.[17]

MEG Processing

Brainstorm software was used to process MEG data. Trials were discarded if they contained eyeblink artifacts, or contained out-of-range activity in any of the sensors (0.1-8000 fT). Three participants exhibited sensor activity consistently out of range, so this metric was not applied to their data as it was not a good marker of abnormal trials. After excluding bad trials, there were at least 375 good trials for every stimulus type for every participant. Data were subsampled as needed to ensure the same number of trials per condition were used in the analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

MEG Decoding

Decoding was performed using the Neural Decoding Toolbox (NDT).[66] We used the maximum correlation coefficient classifier (*MaxCorr*) in the NDT to train classifiers to associate patterns of MEG activity across the sensors with the visual stimuli presented. This classifier computes the mean population vector for sets of trials belonging to each class in the training data and calculates the Pearson's correlation coefficient between those vectors and the test points. The class with the highest correlation is the classifier's prediction. The main conclusions were replicated when using linear support vector machine classifiers. The classifiers were tested using held-out data—i.e., data that was not used in training. Data from both magnetometers and gradiometers were used in the analysis, and data for each sensor was averaged into 5-ms non-overlapping bins from 200 ms before stimulus onset to 600 ms after stimulus onset.

Custom MATLAB code was used to format MEG data preprocessed in Brainstorm for use in the NDT and to combine the two data-collection sessions for each participant. Decoding was performed independently for each participant, and at each time point (5 ms time bins). For each decoding problem, at each time point, the 375 trials for each stimulus condition were divided into 5 sets of 75 trials. Within each set, the 75 trials were averaged together. This process generated 5 cross-validation splits: the classifier was trained on four of these sets, and tested on one of them, and the procedure was repeated five times so that each set was the test set once. This entire procedure was repeated 50 times, and decoding accuracies reported are the average accuracies across these 50 decoding "runs." This procedure ensured that the same data were never used for both training and testing, and it also ensured the same number of trials was used for every decoding problem. The details of the cross-validation procedure, such as the number of cross-validation splits, were determined during the pilot experiments to be those that yielded a high signal-to-noise ratio (SNR) and high decoding accuracy in both participants on the stimulus identity problem.

On each run, both the training and test data were z-scored using the mean and standard deviation over all time of the training data. Following others, we adopted a de-noising method that involved selecting for analysis data from the most informative sensors;[80] we chose the 25 sensors in the training data whose activity co-varied most significantly with the training labels. These sensors were identified as those with the lowest p values from an F-test generated through an analysis of variance (ANOVA); the same sensors were then used for both training and testing. The sensor selection was specific for each participant. The sensors chosen tended to be at the back of the head. Analyses using all channels, rather than selecting only 25, yielded similar results.

Chance classification performance was 1/8 for all analyses except when comparing classification performance using data elicited by hues versus words, in which case classification was binary (chance equal to 50%). For each problem, a classifier was trained and tested in 5ms bins from time t = 200ms before stimulus onset to t = 600ms after stimulus onset. The classifiers' performances were generated through a bootstrapping procedure. First, the problems were evaluated for each participant (resulting in 18 independent decoding time courses). Then, for each unique problem, we averaged the decoding time courses across participants. The gray shading in Figure 2A shows the standard error of the bootstrap mean. To control for individual differences in the absolute time of peak decoding, analyses in Figures 2C, 2E, 2F, and 3 were performed relative to the peak decoding time determined separately for each participant (peak decoding is at 0ms in the figures). For consistency across participants, we restricted the analysis to peak decoding times corresponding to decoding of stimulus onset. Three of the 18 participants showed overall peak decoding times corresponding to decoding of stimulus offset (170-260ms); for these three subjects, the time to peak was constrained to be within the onset time window (0-170ms). The main conclusions are not affected if the analyses were done using absolute peak decoding times. Onset of decoding was defined where decoding accuracy first became significantly above chance for five consecutive 5ms time bins. A permutation test was used to determine which time bins contained significant decoding, and the resulting p values were FDR corrected. For the cross-temporal generalization analysis, each classifier trained using data obtained at each time bin was tested using data obtained at every 5 ms time bin from −200 to 600 ms after stimulus onset (for Figure 5) or from −200 to +1200 (for Figure 4) creating a 2-dimensional matrix of decoding results.

Decoding analyses were also performed using eye tracking data collected during the MEG sessions. Two analyses were conducted: one using pupil diameter and one using eye position (Figure 2A). All parameters were identical to the MEG analysis except for the number of input features to the classifier. Rather than MEG sensors, the classifier used either the diameters of the two pupils (two features) or the xy coordinates of the positions of the two eyes (four features).

RSA and MDS Analysis

To examine the similarities between neural representations of colors, we obtained the results using 'forced-error' classifier tests. In this representational similarity analysis (RSA), classifiers were trained and tested as in the other analyses, but the classifier was prevented from choosing the true stimulus. For each problem, the classifier returned the stimulus that elicited the most similar pattern of

neural activity as elicited by the correct stimulus. This approach was designed not only to yield more data about dissimilarities between stimuli (which could be obtained just by analyzing the errors in the classifier), but also because it is directly analogous to the similarity matching tasks used in behavioral experiments aimed at recovering perceptual similarity among different colors. The dissimilarity matrix was then used in multidimensional scaling (MDS) analysis,[67] to uncover the geometry of the neural representation projected onto two and three dimensions. First, we constructed a confusion matrix of the errors the classifier made for each stimulus, for each individual participant (each column of which was normalized to sum to 1). Then, this matrix was averaged across the diagonal (for instance the instances when light pink was mistaken for light orange were averaged with the instances when light orange was mistaken for light pink) and normalized within each subject so that the largest percentage was equal to 1 and the smallest equal to 0. Finally, these values (measures of confusion) were subtracted from 1 to yield a representational dissimilarity matrix for each participant. These RDMs were averaged together and normalized 0-1. The RDMs of each participant were correlated to three model RDMs (Figures 3A–3C). Gray checks in the models were not included in the analysis (Figures 3B–3C). Significance of the correlation at each time bin was determined using a Wilcoxon signed-rank test, and significance lines were drawn where the correlation was significant for at least 5 consecutive 5ms time bins. The MDS panels were generated using the stress metric, constrained by two dimension (Figure 3D) or three dimensions (Figures S2 and S3; inset Figure 3D).