NeuroImage 274 (2023) 120149

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Categorical working memory codes in human visual cortex ☆

Chang Yan [a,1], Thomas B. Christophel [a,b,1,*], Carsten Allefeld [c], John-Dylan Haynes [a,b,d,e,f]

[a] Bernstein Center for Computational Neuroscience and Berlin Center for Advanced Neuroimaging and Clinic for Neurology, Charité Universitätsmedizin, corporate member of Freie Universität Berlin, Humboldt Universität zu Berlin, and Berlin Institute of Health, Berlin, Philippstraße 13, Haus 6, 10115, Germany
[b] Department of Psychology, Humboldt Universität zu Berlin, Rudower Chaussee 18, Berlin, 12489, Germany
[c] Department of Psychology, City, University of London, London EC1V 0HB, United Kingdom
[d] Berlin School of Mind and Brain, Humboldt Universität, Berlin, Luisenstraße 56, Haus 1, Berlin, 10099, Germany
[e] Cluster of Excellence NeuroCure, Charité Universitätsmedizin, corporate member of Freie Universität Berlin, Humboldt Universität zu Berlin, and Berlin Institute of Health, Berlin, Charitéplatz 1, Hufelandweg 14, Berlin, 10117, Germany
[f] SFB 940 Volition and Cognitive Control, Technische Universität Dresden, Zellescher Weg 17, 01069 Dresden, Germany

ARTICLE INFO

Keywords:
Color
Working memory
Categorical representation
Multivariate
Encoding modeling
fMRI
V1
V4
VO1

ABSTRACT

Working memory contents are represented in neural activity patterns across multiple regions of the cortical hierarchy. A division of labor has been proposed where more anterior regions harbor increasingly abstract and categorical representations while the most detailed representations are held in primary sensory cortices. Here, using fMRI and multivariate encoding modeling, we demonstrate that for color stimuli categorical codes are already present at the level of extrastriate visual cortex (V4 and VO1), even when subjects are neither implicitly nor explicitly encouraged to categorize the stimuli. Importantly, this categorical coding was observed during working memory, but not during perception. Thus, visual working memory is likely to rely at least in part on categorical representations.

Significance statement: Working memory is the representational basis for human cognition. Recent work has demonstrated that numerous regions across the human brain can represent the contents of working memory. We use fMRI brain scanning and machine learning methods to demonstrate that different regions can represent the same content differently during working memory. Reading out the neural codes used to store working memory contents, we show that already in sensory cortex, areas V4 and VO1 represent color in a categorical format rather than a purely sensory fashion. Thereby, we provide a better understanding of how different regions of the brain might serve working memory and cognition.

1. Introduction

The human mind has the ability to temporarily store sensory information to guide decision making and behavior (Baddeley, 1986). Information about memorized sensory contents has been found in activity patterns across numerous cortical regions (Christophel et al., 2017). Importantly, the same memorized content can be represented in more than one region (Christophel et al., 2012; Christophel et al., 2018a; Dotson et al., 2018; Ester et al., 2015; Hernández et al., 2010; Kumar et al., 2016; Sprague et al., 2014), even at the same time (Salazar et al., 2012). While these multiple representations might be simply redundant, it has been suggested that there could be a division of labor such that early sensory regions encode low-level sensory details whereas more higher-level regions represent increasingly abstract and categorical properties of memorized stimuli (Christophel et al., 2017; Fuster, 1997). While prior work has emphasized the precise nature of early sensory representations (Ester et al., 2013), to date, it has remained unclear at which stage the categorical nature of memory representations begins to emerge.

Color stimuli have long been used to study the capacity and precision of visual working memory (Awh et al., 2007; Bays et al., 2009; Buschman et al., 2011; Luck and Vogel, 1997; Wilken and Ma, 2004). Importantly, color stimuli exhibit both continuous and categorical properties. For example, colors are perceived as a continuum but they are also readily grouped into basic color categories (Lindsey and Brown, 2006; Loreto et al., 2012), even when patients are incapable of naming them (Siuda-Krzywicka et al., 2019). Recent behavioral work has suggested that performance during continuous color recall could be explained by a dual content model that combines categorical and continuous (non-categorical) components (Bae et al., 2015; Panichello et al.,

☆ Color should be used for figures in print.

* Corresponding author.
E-mail address: tbchristophel@gmail.com (T.B. Christophel).

[1] These authors contributed equally to the current study

https://doi.org/10.1016/j.neuroimage.2023.120149.
Received 28 July 2022; Received in revised form 24 April 2023; Accepted 29 April 2023
Available online 30 April 2023.

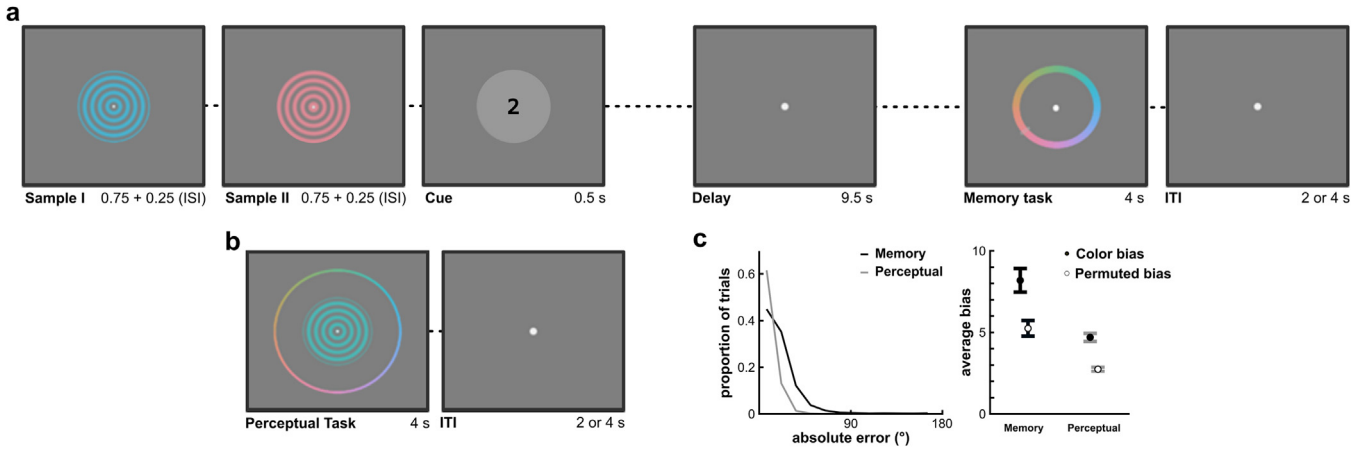

Fig. 1. Experimental design and stimuli. (a) In the working memory task, subjects were presented with two color samples followed by a retro-cue ('1′ or '2′). The retro-cue indicated which of the two items had to be recalled by clicking on the respective color on a color wheel after the delay (memory task). **(b)** In the perceptual task, the sample and the color wheel were shown at the same time minimizing mnemonic demands. **(c)** Subjects made larger errors in the memory task (absolute error in degrees of the color-wheel, collapsed across subjects) and showed more consistent biases in color reports for individual hues (bias defined as absolute error for each color, averaged and compared to biases for permuted color labels, see Methods for details).

2019). Such a dual content model reliably predicted color reproduction that was consistently biased away from the memorized color.

Here we used fMRI and multivariate encoding models to assess whether brain representations of memorized colors in different visual brain areas are better explained by categorical neural codes than by continuous coding. To this end, we scanned 10 healthy participants in multiple MRI sessions (4 each). They performed a conventional color working memory task requiring subjects to recall a remembered color as accurately as possible and indicate their choice with a continuous color wheel. To clearly separate categorical biases in mnemonic activity from perceptual categorization effects, we used separate memory and perceptual tasks minimizing potential overlap in the hemodynamic responses to perceptual and mnemonic activity. Thus, subjects either recalled the colors immediately (undelayed 'perceptual' task, see Fig. 1**b**) or after a delay (delayed 'memory' task, see Fig. 1**a**). To closely capture the neural activity patterns encoding colors of different hues, we sampled colors evenly from a calibrated color space.

2. Materials and methods

2.1. Participants

Ten right-handed healthy German native speakers (aged 18–35 years; mean age: 27, SEM ± 1.13; 9 female) with normal or corrected-to-normal vision and no color blindness participated in the study. The sample size and the number of repetitions per task was chosen based on previous studies using similar analyses techniques to study perceptual color representations as well as working memory representations (Brouwer and Heeger, 2009; Rademaker et al., 2019), and was considerably increased. We decided to recruit a small subject number with multiple sessions per subject, instead of a large number of subjects (Cosgrove et al., 2007). This study was granted ethical approval by the local ethics committee and all subjects gave informed consent.

2.2. Experimental design

Each subject completed five sessions of experiments, including three 2-h fMRI sessions with 16 runs (50 trials/run) for a delayed estimation task ('memory task'), one 2-h fMRI session with 14 to 16 runs (50 trials/run) for an undelayed estimation task ('perceptual task') both using color as stimulus material (see Fig. 1**ab**). The third and last session was one 90-min behavioral session for the color categorization tasks (see Fig. 2**ab**). These five sessions were conducted on different days, but within the same month. After the last fMRI session, participants also completed a 2-page questionnaire regarding their strategies for completing the working memory task. All experimental tasks were coded using PsychToolbox-3 (http://psychtoolbox.org/) and MATLAB 2014b (MathWorks, Natick, MA).

2.2.1. Memory task

In the delayed estimation task ('memory task'), subjects memorized a sample hue during a delay period and then reported the memorized color on a randomly rotated color wheel. A trial started with the sequential presentation of two color samples in the middle of the screen, followed by a retro-cue (Sperling, 1960) (either '1′ or '2′) at the center of a light gray circle (see Fig. 1**a**). The sample stimuli were concentric sinusoidal gratings within a circular aperture changing from the central gray point to the sample color, which drifted at a constant speed in a random direction: either inward or outward (Brouwer and Heeger, 2009). A retro-cue informed subjects which of the two sample stimuli should be memorized for the rest of the trial ('1′ or '2′). The retro-cue was followed by the presentation of a blank screen (with only the fixation point) for 9.5 s, resulting in an overall delay of 10 s for memorization of the cued stimulus. Then a color wheel included all 50 color samples was presented in the center of the screen. Subjects were asked to indicate on the color wheel which sample they had memorized within 4 s. For this, they scrolled with an MRI compatible trackball from the screen center (where the cursor was a white dot) onto the color wheel (where the cursor changed to a white rectangular box), and by clicking a button to confirm their choice. Once the selection was confirmed, both the color wheel and the response remained on the screen until the end of 4 s. The color wheel was rotated by random degrees in each trial, thus avoiding confounding motor preparation with the reported color. Subjects were required to fixate throughout the trial.

The duration of one trial was either 18 s or 20 s, including an inter-trial interval (ITI) of 2 or 4 s (on average ITI = 3 s). A run was comprised of 50 trials in random order, with each of the 50 sample stimuli presented once as the cued stimulus and the not cued stimulus (fully randomized from each other). Three fMRI scanning sessions resulted in altogether 16 runs and 800 trials for the delayed estimation task per subject. Before the first scanning session, subjects were trained for half an hour with feedback on their responses.

2.2.2. Perceptual task

In the undelayed estimation task ('perceptual task'), subjects reported a seen color on a concurrently presented color wheel. A trial

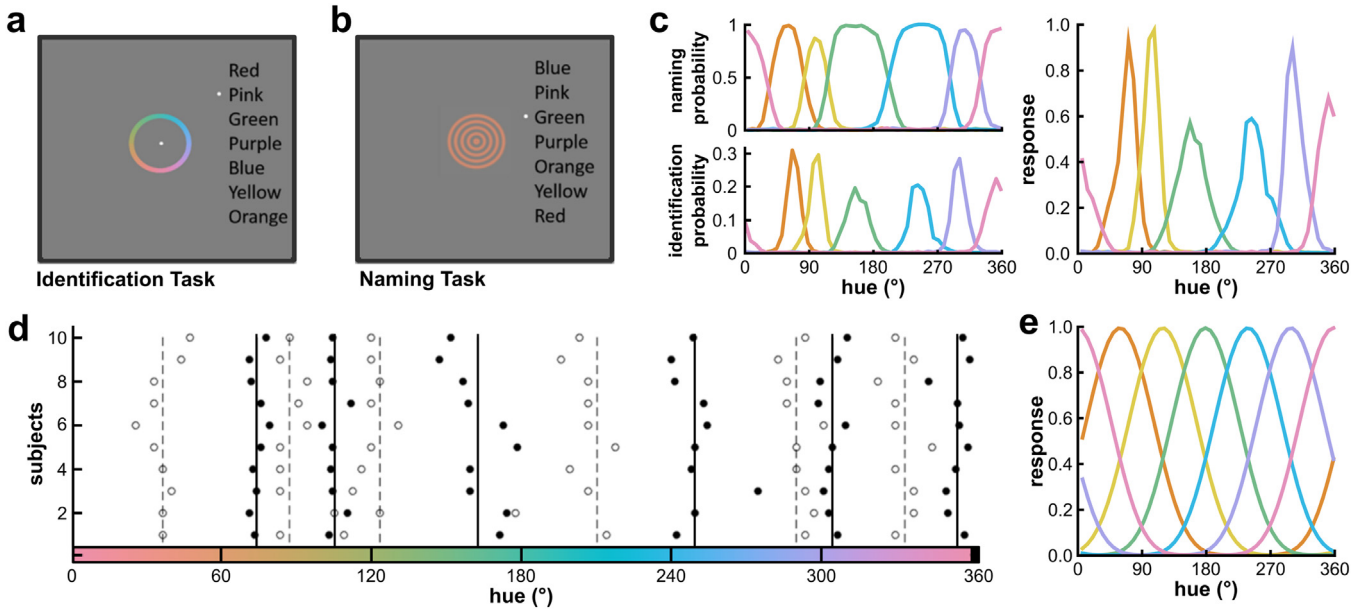

Fig. 2. Behavioral categorization tasks and resulting encoding models. We used two category-based tasks performed in a separate session post-MRI where subjects **(a)** identified prototypical colors on a color wheel and **(b)** named presented colors.**(c)** We combined naming and identification frequencies for individual hues to create a tuning model for voxels selective to six color categories by averaging the two distributions **(as in [22]). (d)** This model allows to estimate the boundaries between color categories (open circles, dashed lines) and the prototypical exemplars for each color category (closed circles, solid lines). **(e)** For comparison, we used a standard uniform model typically thought to capture the sampling of sensory units by an fMRI voxel.

began with the presentation of a sample stimulus in the middle of the screen (see Fig. 1**b**). 500 ms after sample onset, the color wheel started to fade into view within a period of 350 ms. The color wheel was faded-in to minimize interference with the subjects' perception of the color sample. The sample was presented for 4 s, the response remained on the screen after subjects selected the seen sample on the color wheel. The next trial started after an inter-trial interval of 2 or 4 s (on average ITI = 3 s). A trial was thus either 6 s or 8 s (on average 7 s), and a run consisted of 50 or 100 trials. Altogether 700 to 800 trials were conducted for the undelayed estimation task per subject.

2.2.3. *Category naming and identification tasks*

A pair of behavioral categorical tasks, color naming and identification, were performed in order to delineate the properties of color categories in our sample. The tasks were conducted in a dark behavioral lab using a keyboard and a mouse, after the completion of all fMRI sessions. Subjects had no time pressure as the next trial only started after they completed the current trial.

In the color naming task (Fig. 2**b**), a list of seven common color names including 'blue', 'pink', 'green', 'purple', 'orange', 'yellow' and 'red' was shown next to a sample color. These chromatic color terms were selected based on Berlin and Kay's eight basic color categories (Berlin and Kay, 1969) but 'brown' was excluded (see prior work, Bae et al., 2015). Subjects were asked to select the term that best described the color stimulus by pressing the up or down button on the keyboard, and to confirm their choice by pressing enter. The order of the terms as well as the initial position of the cursor were randomized in each trial to minimize position bias. Six subjects completed 12 trials for each of the 50 color stimuli, while four subjects evaluated each stimulus 9 or 6 times (due to time constraints of the behavioral session).

In the color identification task (Fig. 2**a**), subjects were required to mark the color wheel to identify the prototypical exemplar for of each of the seven color terms (see above). By pressing the left button of the mouse, they could confirm the color selection. The color wheel was rotated by random degrees in each trial to prevent association between the position and the color. Six subjects completed 90 identification trials for each category term, while four subjects evaluated each term 60 times.

2.3. *Stimuli*

We used a set of 50 color samples taken from a circular color space with constant lightness (CIE LAB; center: $a^* = 0$, $b^* = 0$; radius = 38; L^*=70; Fig. 1**c**). Using a large number of different colors allowed us to finely sample variations in neural coding for stimuli in this circular space. A spectroradiometer (JETI spectraval 1501) was employed to measure $L^*a^*b^*$ values of each of the 50 generated colors, and to calibrate these parameters on different screens (the MRI monitor for the MRI session and the computer screen for the behavioral session). More specifically, we first calibrated the background gray color (used as the reference white point) to approximate a XYZ ratio of 1:1:1. Then, each color stimulus was measured and changed in multiple iterations to minimize the discrepancy to the chosen $L^*a^*b^*$ values.

2.4. *Data acquisition*

MRI data were acquired on a 12-channel Siemens 3 Tesla TIM-Trio scanner at the Berlin Center for Advanced Neuroimaging (BCAN). At the beginning of each scanning session, a high-resolution T1-weighted magnetization-prepared rapid gradient echo (MPRAGE) anatomical volume was collected (192 sagittal slices; repetition time TR = 1900 ms; echo time TE = 2.52 ms; flip angle = 9°; FOV = 256 mm). For acquisition of functional BOLD imaging, T2*-weighted echo planar images (EPI; 32 contiguous slices; TR = 2 s; TE = 30 ms; voxel size = 3 × 3 × 3 mm; matrix size = 64 × 64 × 32; slice gap = 0.6 mm; descending order; flip angle = 90°; FOV = 192 mm) were recorded covering the whole neocortex. Every trial was time-locked to the start of an EPI acquisition. For the memory task, 478 EPI scans were collected per run, and altogether 7648 scans were acquired over 16 runs per subject. For the perceptual task, data was acquired either in single (50 trials, 175 scans) or double (100 trials, 350 scans) runs. Overall, 2450 to 2800 functional scans were recorded per subject.

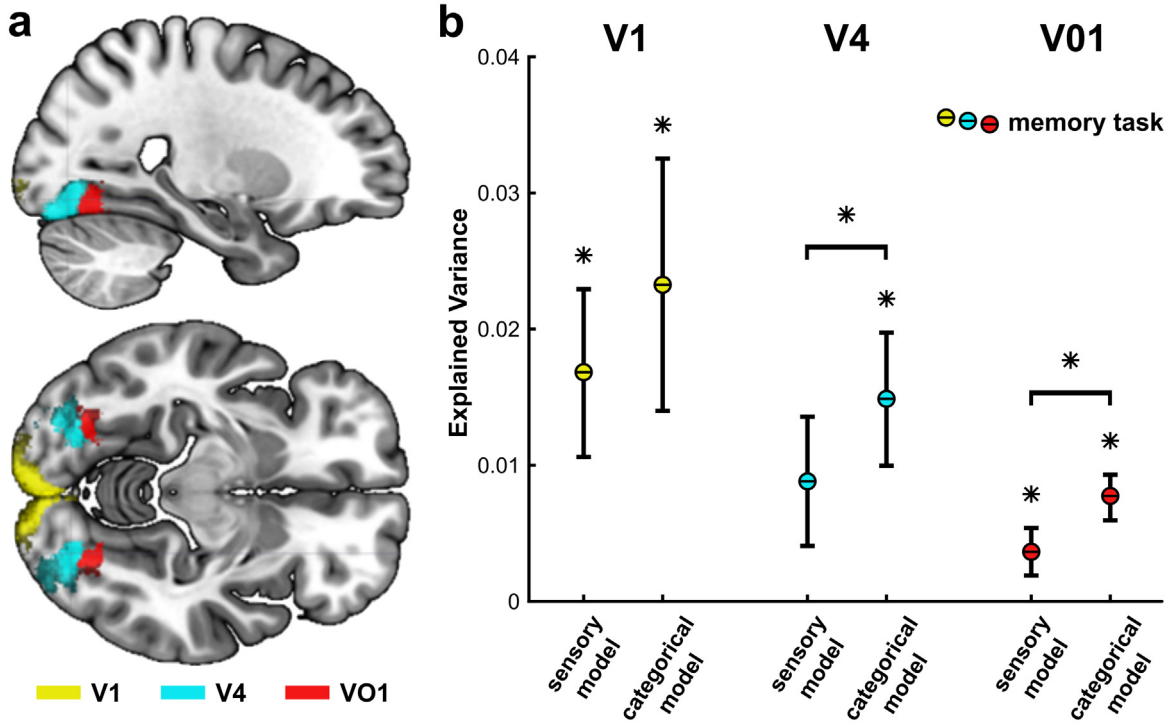

Fig. 3. Categorical working memory codes in visual cortex. (a) We focused our analyses on anatomically defined regions of interest for V1, V4 and VO1. **(b)** Explained variance in fMRI voxel activity patterns for different encoding models (sensory and categorical) during the memory task. Regions are color coded; asterisks indicate significance (bootstrapped confidence intervals with $p < 0.05$; multi-comparison corrected); error bars indicate SEM.

2.5. Behavioral analyses

To quantify the precision of color recall, we calculated the absolute error across trials separately for each condition (memory task and perceptual task) expressed in degrees of the 360° color-wheel. We also wanted to know whether subjects' color reproductions were more biased during working memory as compared to a perceptual task (Bae et al., 2015). For this, bias in color reports was quantified by averaging recall errors for all repetitions of a given color (14–16 repetitions) and then averaging the absolute value of this individual bias across all colors. This bias metric, however, will have a non-zero value even if no consistent biases exist in the data depending on the overall error in the task. To control for this effect of random (color independent) recall errors on this bias measure, we permuted the color labels for each trial 1000 times and compared the resulting bias estimates against the biases computed from the real labels.

Data from the naming and the identification task was used to identify boundaries and prototypical colors for common color categories (for illustrative purposes, see Fig. 2**d**). Naming data was minimally smoothed (Gaussian smoothing window, size 2, $\sigma = 0.5$) to minimize noise when determining boundaries between categories.

2.6. Anatomical regions of interest

We focused our analyses on BOLD data from three regions of interest (ROIs) in visual cortex: V1, V4, and VO1. These ROIs (Fig. 3**a**) were delineated based on high-resolution anatomical probabilistic maps (Wang et al., 2015). These high-resolution probability maps were processed to obtain binary maps for every individual subject. First, the maps were deformed into the brain space of individual subjects using (inverse) normalization parameters obtained using unified segmentation. Then, the maps on the left and right hemispheres were collapsed and dorsal and ventral components were combined. We applied a mutual exclusion rule for all available probabilistic maps (Wang et al., 2015), such that every voxel could only be part of one ROI by selecting the ROI label with the highest probability. Finally, we threshold the resulting subject-level maps to exclude voxels with a probability lower than 10% to obtain a binary ROI map. For a post-hoc analysis focusing on frontal regions we created regions of interest using MNI coordinates from regions found to carry information about sensory stimuli in prior work (Ester et al., 2015; Yan et al., 2021). In particular we used spheres with a radius of 11 mm around left ventrolateral prefrontal cortex (MNI = [−37 30 −2], Ester et al., 2015; average coordinate in Table 2), left premotor cortex (MNI = [−46 10 48]; Yan et al., 2021), and Broca's area cortex (MNI = [−56 34 −4]; Yan et al., 2021) and transformed the resulting regions into single-subject space.

2.7. fMRI preprocessing

All fMRI analysis was conducted using SPM12 (Friston et al., 1994), cvMANOVA (Allefeld and Haynes, 2014), and Matlab 2014b (Mathworks, Natick, MA). The acquired images were first converted from DICOM format to a SPM compatible format of NIfTI. Next, all functional images belonging to one subject were realigned and resliced to correct for head movement within and between runs. Then, the anatomical image was coregistered to the first functional image and subjected to unified segmentation (for inverse normalization).

2.8. Sensory and categorical encoding models

To estimate color-selectivity from spatially scattered and distinct response patterns of a population of voxels, we use two distinct encoding models: one sensory continuous and one categorical model (see Fig. 2c, e). Encoding models capture the pattern of selectivity a voxel can be characterized as the weighted sum of a set of color-selective channels analogous to a neuron's tuning curve. The *sensory* encoding model was characterized by six half-wave rectified cosine functions evenly distributed over the circular color space and raised to the power of six (Fig. 2e). Such encoding models were used in prior work to model the selective neural response to orientation, spatial location

and color (Brouwer and Heeger, 2009; Edward F. Ester et al., 2015, 2013; Sprague and Serences, 2013) and are intended to approximate single-unit tuning functions of *sensory* cortical neurons (Brouwer and Heeger, 2009; Ester et al., 2013).

To model *categorical* neural representation, we developed a novel type of basis function of color (Fig. 2c) using empirical color categorization data (Fig. 2ab).

Six categorical basis functions captured the boundaries and prototypical colors of 'blue', 'pink', 'green', 'purple', 'orange', and 'yellow'. To demarcate the boundaries of selectivity categories, we used the category naming data, and the category identification data was used to identify the prototypical exemplars of each category. The two corresponding probability distributions combining data from all subjects (Fig. 2c, left) were normalized, so that the sum probability of each category equaled one. Then, we averaged the two probability distributions to create an encoding model capturing both the boundaries and the prototypical exemplars of each category. The resulting basis functions were normalized by dividing them by the highest value among all six channels.

Notably, this categorical encoding model does not require category-selective neuronal populations to exhibit an all-or-none response to any given stimulus. This is motivated by the behavioral data indicating that color categorization is probabilistic with the same hue being assigned different color categories in different trials in the naming task. Further, we anticipated that categorically color-selective voxels respond strongest to prototypical exemplars of a given color category. It is important to note that this graded categorical model predicts that prototypical members of a color category evoke overall more univariate activation than atypical category members. This property is intended to incorporate the uncertainty of the categorization of particular exemplars. Cat and dog selective neurons in ITC and PFC, for example, show a graded response to more or less prototypical exemplars (Freedman et al., 2003). Finally, fitting the six basis functions simultaneously allows any voxel to have positive weights for multiple color categories as it might contain neurons selective for multiple categories.

Thus, we created two distinct encoding models: (1) A *sensory* model used in prior work to resemble the tuning of sensory neurons while carrying no information about the delineations of common color categories, and (2) a categorical encoding model informed by empirical categorization data.

In a post-hoc analysis, we also created a third model variant to interrogate the particular shape of the categorical encoding model used in the main analyses. For this we used the boundaries between color categories derived from the color naming data to build a set of boxcar shaped regressors representing the six color categories. In this model, neural activity patterns are assumed to be identical for both prototypical and atypical colors of a category. We also interrogated the specific shape of the sensory model. For this we used a 'steerable' encoding model using one sine and one cosine function in hue space. Sine and cosine functions with an arbitrary phase allow for the reconstruction of sines and cosines of any other phase, allowing them to capture the entire hue space regardless of the phase and the corresponding 'channel centers' with only two regressors (i.e. making the model 'steerable', see Brouwer and Heeger, 2009; Freeman and Adelson, 1991). Notably, this steerable model differs from previous implementations (e.g. Brouwer and Heeger, 2009) as it includes both positive and negative regressor values and is used across the full cycle of sin and cos. As in all other model, this steerable model included a constant term regressor.

It is important to note here, that no a-priori defined encoding model can be expected to perfectly fit neural data in a particular area. The behavioral measures obtained to create the categorical model here are influenced by several processing stages, including sampling from (potentially multiple) categorical representations of color, as well as verbal processes involved in the naming and identification tasks and are unlikely to perfectly resemble categorical tuning in any particular region. Reversely, the continuous regressors intended to approximate single-unit tuning functions of sensory cortical neurons are unlikely to resemble the exact tuning of any set of sensory neurons (or the resulting voxel-wise tuning in fMRI). This is why we elected to use models used in prior work (Bae et al., 2015; Brouwer and Heeger, 2009; Ester et al., 2015, 2013; Sprague and Serences, 2013) as a means to limit the search space. Please note that many previous studies have employed similar idealized encoding models because the true model is unknown. While we take inspiration from this prior work the modeling approach employed here (i.e. predicting the neural data) differs approaches used in this previous work (i.e. predicting a stimulus or response), meaning we do not intend to fully emulate it.

2.9. Multivariate pattern analysis

To test which of these models best explained mnemonic activity patterns during the working memory delay, we combined these two encoding models with a recently-developed form of multivariate pattern analysis (MVPA), cross-validated multivariate analysis of variance (cvMANOVA; Allefeld and Haynes, 2014). We used cvMANOVA to directly estimate and compare the explained variance of the two encoding models. An alternative approach could be comparing the similarity (or dissimilarity) between remembered items to similarities predicted by the two encoding models (Kriegeskorte et al., 2008). For a given pair of representations with a fixed similarity, however, there can exist a manifold of possible underlying encoding patterns meaning that the same pattern of similarities can arise due to starkly different encoding schemes. Directly estimating the fit of the acquired neural data avoids this source of ambiguity and can be expected to be a powerful tool.

The analysis was performed on a set of selected voxels within three regions of interest (ROIs): V1, V4, VO1 (see Fig. 3a). For this, we first estimated parameters (i.e., betas) for multivariate generalized linear models (MGLM) separately for each condition (memory and perceptual) and encoding model (sensory and categorical) which modelled sample colors as a set of six parametric modulations representing the six basis functions per model.

For the memory task, we used five finite impulse response (FIR) regressors to represent the 10 s delay-period (5 fMRI scans at a TR of 2 s). The design matrix modelled the 478 scans per run using 36 regressors (7 stimulus-based regressors [1 constant and 6 basis functions] x 5 FIR bins + 1 run-wise constant). Separate design matrices and MGLMs used basis functions from the *sensory* and the *categorical* encoding model.

For the perceptual task, but the 4 s stimulus presentation was represented by a canonical hemodynamic response function (HRF, duration = 4 s), which was time-locked to the stimulus presentation's onset. The design matrix captured each run (either 175 or 350 scans) using 8 regressors (7 stimulus-based regressors [1 constant and 6 basis functions] x 1 HRF + 1 run-wise constant). Again, separate design matrices and MGLMs used basis functions from the *sensory* and the *categorical* encoding model. Parameter estimates for all models were estimated using standard SPM parameters, but parametric modulations were not orthogonalized and serial correlations corrections were omitted.

Next, we estimated the variance explained by these models by contrasting each neighboring pair of basis functions (BF 1 vs BF 2; BF 2 vs BF3; BF 3 vs BF 4…) separately for each time point (for the memory task). We elected to contrast data for each time-point independently to allow the contrast to account for variations in neural code over time but focused on estimates of explained variance that are averaged across the delay to increase power. For the memory task, the overall contrast matrix for a given run was comprised of 35 columns representing 35 regressors (six BF-based and one stimulus-based regressors, each in five FIRs bins) and 25 rows representing 25 contrasts (five contrasts between six BF-based regressors, each in five FIR bins). For the perceptual task, the contrast matrix for a given run had 7 columns representing 7 regressors (six BF-based and one stimulus-based regressors, each in one HRF bin) and 5 rows representing 5 contrasts (five contrasts between six BF-based regressors). The null hypothesis, here, is that in a given set of voxels there are no differences in the parameter estimates for the

six basis functions in a given model. Rejecting this null indicates that this subset of data carries information about sample color in a given trial. The resulting pattern distinctness D (Allefeld and Haynes, 2014; Christophel et al., 2018b; Yan et al., 2021) reflects the variance of the neural data explained by the respective model, cross-validated across runs. Here, if different colors elicit the same multivariate response, D would on average be 0, while different responses to different colors that are captured by a given encoding model would lead to an average D larger than 0. To assess statistical significance against chance-level ($D = 0$) and to compare models against each other as well as model-by-task interactions, we used a nonparametric bootstrapping testing group effects by random resampling 10^5 times (Bickel and Freedman, 1981; Efron, 1979; Singh, 1981). We corrected the resulting confidence intervals for the multiple comparisons in the three different ROIs using Bonferroni correction (resulting in an effective confidence interval of 98.33%).

To validate our encoding model approach using cvMANOVA, we performed two separate simulations using the two encoding models to generate artificial data. We assigned a set of 6 random weights to each simulated voxel (100 voxels overall) corresponding to the six basis functions in a given encoding model. The prototypical pattern of neural activity for a given color was computed by weighted averaging of the responses of the six basis functions for this specific color weighted using voxel-wise weights for each basis function. For each simulated subject ($N = 10$), we generated a dataset of 16 runs and 50 trials in each run using the 50 colors as stimuli, using different random weights for each subject and voxel, and adding random Gaussian noise in each trial. We then included 300 data points containing only random Gaussian noise that did not contain any color representation to simulate time between trials. The noise component was weighted using a factor (19 even steps, ranging from 1 to 73 in different iterations) to simulate different signal-to-noise ratios. The two resulting datasets were analyzed as in the main analysis and the overall procedure was repeated 10.000 times. We report the proportion of these 10.000 Iterations where a given encoding model (i.e. the categorical model) explained significantly more variance than the other model (i.e. the sensory model) for data generated the two models. In addition, we simulated data sets that contained no color representations five times.

3. Results

Subjects made larger errors in the memory task (mean absolute error = 17.47° ±1.52° SEM in degrees of the color-wheel, see Fig. 1c) than in the perceptual task (mean absolute error = 9.8°±0.39° SEM, Wilcoxon signed rank test, $p = 0.002$), and showed larger categorical biases independently of the overall effect of the errors (see Fig. 1c, Wilcoxon signed rank test, $p = 0.02$). Subject indicated the use of both visual and non-visual encoding strategies during the post-experimental questionnaire (see Supplementary Fig. 1).

In a behavioral session after the fMRI experiments, we used two separate category-based behavioral tasks to obtain a categorical model of color representation. Following prior behavioral work (Bae et al., 2015), subjects performed two tasks: In the *color identification* task they were given a color name and asked to identify that color on a continuous color wheel (see Fig. 2a). In the *color naming* task subjects assigned a color name to a continuous color (see Fig. 2b). We used 7 color names for these tasks but only six were consistently used by the participants in the naming task. As in prior work (Bae et al., 2015), data for the seventh, unused color name ('red') was discarded. The resulting behavioral data allow to assess two properties of color representation: the boundaries between color names and the prototypical exemplars for the six most commonly used color categories (see Fig. 2d and Methods for details). We then averaged the underlying naming and identification distributions to form a simplified categorical encoding model (see Fig. 2c). Such a model has been used in the past to predict behavioral bias in color recall (Bae et al., 2015), but serves here as an approximation of the tuning of category-selective neurons for these color categories responding strongest to the prototypical color and showing a sharp decline towards the boundaries. For comparison we used a standard cosine-shaped continuous sensory (non-categorical) encoding model (see Fig. 2e), versions of which have been used in previous investigations of working memory coding and precision (Brouwer and Heeger, 2009; Edward F. Ester et al., 2015, 2013; Sprague et al., 2014). This sensory model did not consider information about the delineations of common color categories.

We then assessed which of these two models best explained mnemonic representations of color in the brain. Using probabilistic anatomical regions of interest (Wang et al., 2015) (see methods for details), we focused our analyses on regions of the visual cortex known to have representations of color (Fig. 3a; V1, V4 and VO1, Brouwer and Heeger, 2009). Notably, we used a cross-validated form of multivariate analysis of variance (cvMANOVA; Allefeld and Haynes, 2014, see methods for details) to assess which of the two models better explained the underlying data instead of inverting the model to reconstruct the memorized or the reported color (see Rademaker et al., 2019). This is important as we aimed at identifying intermediary representations that not necessarily match either the encoded or the reported color.

Using the sensory model, we found robust representations of the memorized color in V1 and VO1 (Fig. 3b; 95% bootstrapped confidence interval of the explained variance corrected for multiple comparisons; CI^{95} = [0.0017, 0.0292] in V1, CI^{95} = [−0.0038, 0.0181] in V4, CI^{95} = [0.0003, 0.0082] in VO1). When using the categorical model, all three ROIs showed statistically significant information about the target color (CI^{95} = [0.0043, 0.0469] in V1, CI^{95} = [0.0014, 0.0243] in V4, CI^{95} = [0.0041, 0.0124] in VO1). Notably, the two models are inevitably non-orthogonal with respect to each other, meaning that they can be expected to capture shared variance. Hence, we statistically compared which of the two models explained more variance in the underlying data. We found that the categorical color model explained more variance in V4 and VO1 than the sensory model that was not informed by the delineations of common color categories (Fig. 4b; 95% bootstrapped confidence interval corrected for multiple comparisons of the categorical model preference; CI^{95} = [0.0017, 0.0152] in V4, CI^{95} = [0.0012, 0.0065] in VO1). This suggests that during working memory V4 and VO1 encode memoranda in a categorical rather than a sensory format. In contrast, neither model outperformed the other model in V1 (CI^{95} = [−0.009, 0.0212]). The pattern of results stayed the same when we contrasted all five timepoints together, but the model comparison was not significant in V4 (one-sided 95% bootstrapped confidence interval of the categorical model preference; CI^{95} = [−0.0036] in V1, CI^{95} = [−0.0002] in V4, CI^{95} = [0.0006] in VO1). Testing the same effects for any differences across the five time-points in the delay using a nonparametric repeated-measures ANOVAs (Gladwin, 2020) separately for each region we found a main effect of model in both V4 and VO1 (all $p < 0.05$) and no main effects of time and no time-by-model interactions (all $p > 0.1$).

Then, we asked how color was represented when subjects did not have to retain the color samples in memory but could immediately report the presented color. Importantly, subjects used the same method to report the perceived color in the perceptual and the memory task to minimize biases due to different task-goals (Lee et al., 2013). For data from this perceptual task, the sensory model explained significant levels of variance in all three regions (Fig. 4a; CI^{95} = [0.0198, 0.0724] in V1, CI^{95} = [0.0085, 0.0781] in V4, CI^{95} = [0.0013, 0.034] in VO1). In contrast, using the categorical encoding model only V1 and V4 but not VO1 explained above-chance variance (CI^{95} = [0.0202, 0.0783] in V1, CI^{95} = [0.0082, 0.0495] in V4, CI^{95} = [−0.0055, 0.0222] in VO1). Critically, for the perceptual task we found no significant differences in the variance explained by the two models in any region (CI^{95} = [−0.0174, 0.0149] in V1, CI^{95} = [−0.0384, 0.0049] in V4, CI^{95} = [−0.0144, 0.0035] in VO1). Comparing these differences across the two versions of the task (delayed and undelayed) resulted in a significant interaction effect in VO1 (Fig. 4b; 95% bootstrapped

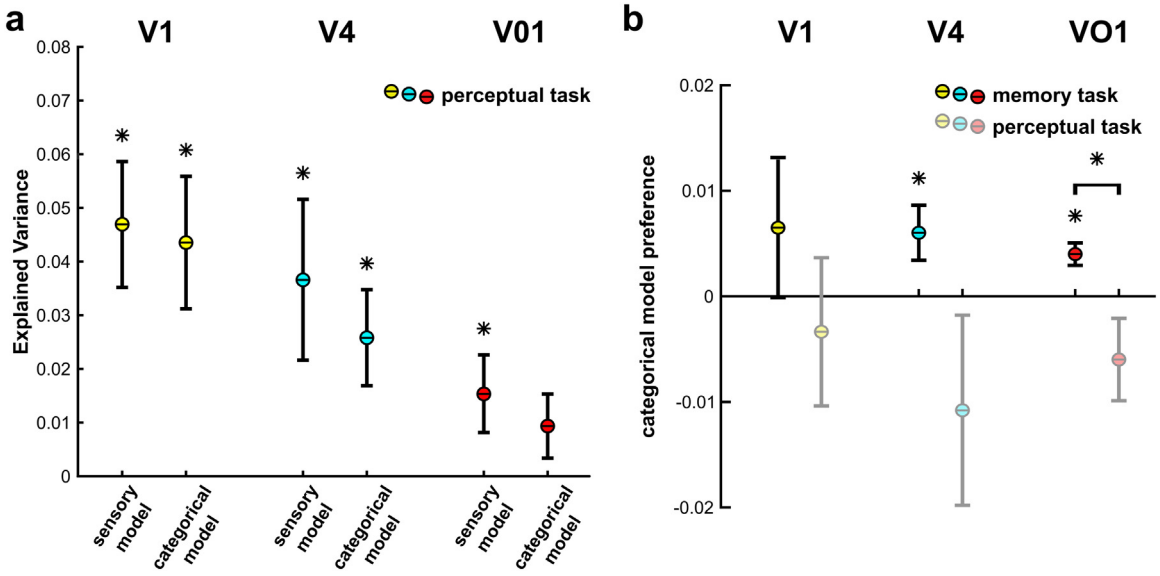

Fig. 4. Color representation during the perceptual task (a) Explained variance in fMRI voxel activity patterns for different encoding models (sensory and categorical) during the perceptual task. Regions are color coded; asterisks indicate significance (bootstrapped confidence intervals with $p < 0.05$; multi-comparison corrected); error bars indicate SEM. (b) Categorical model preference in the three regions during the two tasks. Please note that categorical model preferences in V4 are significantly larger during memory as compared to perception when multiple comparisons corrections are omitted ($p < 0.05$, CI95 = [−0.0377, −0.0011]).

confidence interval corrected for multiple comparisons of the difference in categorical model preference; CI^{95} = [−0.0238, 0.0118] in V1, CI^{95} = [−0.0429, 0.0016] in V4, CI^{95} = [−0.0177, −0.0023] in VO1) indicating that representations in the delayed task were substantially more categorical than in the undelayed task in VO1. To investigate whether the categorical model preference solely relied on the atypical encoding model shape of the categorical model we used a boxcar shaped encoding model relying solely on naming probability data. We found that this boxcar shaped categorical model explained more variance in V4 and VO1 than the sensory model (one-sided 95% bootstrapped confidence interval of the categorical model preference; CI^{95} = [0.0012] in V4, CI^{95} = [0.0001] in VO1). We found no significant difference in V1 (CI^{95} = [−0.0095]). The same pattern of results was found when equating the peaks of the original categorical encoding model to 1 (one-sided 95% bootstrapped confidence interval of the categorical model preference; CI^{95} = [0.0027] in V4, CI^{95} = [0.0024] in VO1, CI^{95} = [−0.0080] in V1). Finally, we tested whether a categorical encoding model created using the naming and identification data of each individual subject to model the fMRI data of that subject was superior to the model combining data from all subjects that was used throughout the manuscript. Testing this subject-specific encoding model we found that the overall pattern of results preserved and no significant differences between the subject-specific and the subject-general model (95% bootstrapped confidence interval corrected for multiple comparisons of the subject-specific model preference; Memory task: CI^{95} = [−0.0119 0.0061] in V1, CI^{95} = [−0.0060 0.0080] in V4, CI^{95} = [−0.0023 0.0044] in VO1; Perceptual task: CI95 = [−0.0069 0.0188] in V1, CI95 = [−0.0258 0.0044] in V4, CI95 = [−0.0065 0.0048] in VO1). We also tested whether a 'steerable' encoding model using sine and cosine functions would be more suitable to investigate working memory data. This model was outperformed by both the sensory model (one-sided 95% bootstrapped confidence interval of the sensory model preference; CI^{95} = [0.0004] in V1; CI^{95} = [0.0010] in V4, CI^{95} = [0.0006] in VO1) and the categorical model (CI^{95} = [0.0044] in V1; CI^{95} = [0.0063] in V4, CI^{95} = [0.0040] in VO1).

In a post-hoc analysis, we further explored whether frontal regions found to represent working memory content in prior work (Ester et al., 2015; Yan et al., 2021, see methods for details) represented memorized colors and demonstrate preferences for the categorical or the sensory model. We found that both left ventrolateral prefrontal cortex and left premotor cortex represented the memorized color when probed with the sensory model (95% bootstrapped confidence interval corrected for multiple comparisons of the explained variance; CI^{95} = [0.0004 0.0185] in L-VLPFC, CI^{95} = [0.0017 0.0162] in left L-PMC) but not when tested with the categorical model (CI^{95} = [−0.0118 0.0235] in L-VLPFC, CI^{95} = [−0.0019 0.0203] in left L-PMC). There were no significant differences between the variance explained by the two models in either area (95% bootstrapped confidence interval corrected for multiple comparisons of the difference in categorical model preference; CI^{95} = [−0.0131 0.0023] in L-VLPFC, CI^{95} = [−0.0061 0.0115] in L-PMC). We found no reliable color representations in the Broca's area ROI using either model (CI^{95} = [−0.0016 0.0069]; CI^{95} = [−0.0043 0.0157]; respectively).

Finally, to validate our encoding model approach, we performed 10.000 iterations of two separate simulations using the two encoding models to generate artificial data. For each iteration, we generated a data set (10 subjects, 16 runs, 50 trials, 100 voxels, see Methods for full details) where either the categorical or the sensory model was the true model underlying simulated neural representations and analyzed the data as in the main analyses reported above. In these analyses, we see a preference for the categorical encoding model when the categorical model was used to generate the data and a preference for the sensory encoding model when the data was based on the sensory model (see Fig. 5). The proportion of iterations with a significant model preference declined as noise increased and data without any simulated neural representations only showed preferences for either model at a rate of 0.05.

4. Discussion

These results suggest that representations of memorized colors are retained by categorical representations already within visual cortex. This suggests that already in visual cortex, different regions retain representations in a differential neural code. Prior work provides evidence than primary sensory cortices are critical for precise representations of orientation (Ester et al., 2013). In concert, these findings seem to indicate that memory storage of an individual item does rely on its representation in multiple concurrent coding schemes.

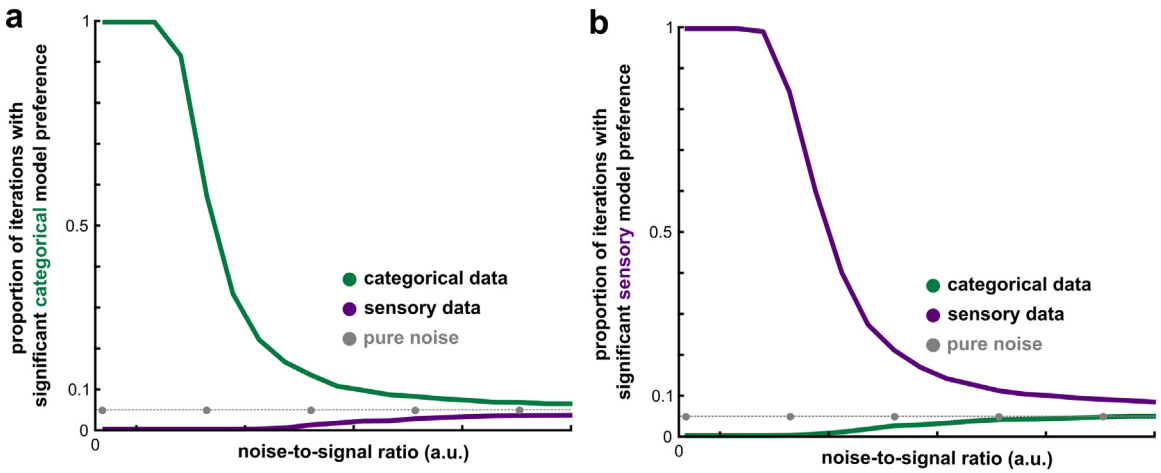

Fig. 5. Validation using simulated data. (a) Proportion of simulation iterations showing a significant categorical model preference for data generated using either the categorical or the sensory or encoding model. **(b)** Proportion of iterations showing a significant sensory model preference.

Importantly, the encoding models used to distinguish between continuous and categorical working memory codes are somewhat correlated (median absolute $r = 0.36$) This becomes apparent when realizing that both models represent the same property (color) and that they are both members of an infinitely large family of models representing this property in different ways. Thus, finding that any given model explains variance is not evidence that the underlying neuronal population uses exactly this neural code. Showing that the categorical model outperforms the sensory model allows us to infer that within this large family of models, the 'true' model is likely to share more properties with the categorical model than with the sensory model. The simulations reported above demonstrate the general feasibility of this approach and the imaging results reported show that model distinctions are possible in real imaging data.

Notably, this preference for categorical tuning is only present during working memory and is absent during an immediate, perceptual task. This suggests that initially more uniform representations during perception morph to become more categorical during the working memory delay. The comparatively short delay and the retro-cuing procedure in the current study prohibits us from directly interrogating changes in categorical preference across time. Thus, two alternative explanations might explain this difference. One possibility is that the affordance to memorize an item for a prolonged period of time results in a more categorical encoding of the stimuli. Alternatively, feedback from more anterior, non-sensory regions might slowly modulate representations during the course of the delay. Both alternatives provide a direct explanation for the differential amount of biasing of behavioral responses in immediate and delayed recall in this study and in prior work (Bae et al., 2015; Panichello et al., 2019).

This demonstrates that the comparison between perceptual and memory representations of colors employed in the current study might include additional effects that could have contributed to the findings. Here, we aimed at closely emulating previous behavioral work demonstrating differential biases between delayed and immediate **recall** (Bae et al., 2015). This required that subjects were tasked to recall the presented color immediately in the perceptual condition (i.e., without any need for memorization) resulting in an overlap of perceptual and task-related signals in the recorded data. Future work might instead choose to compare mnemonic signals to task-irrelevant ('unattended') perceptual signals (see Harrison and Tong 2009), which however in turn entails the caveat that differences in cortical representation might be a result of inattentive processing. A second alternative would be to delay task-execution effectively making the perceptual task a memory task with a shorter delay (or comparing signals early and late in the delay). This second alternatively, however, compares working memory representations to a mixture of perceptual and working memory representations. For this it is worth noting that increased biases in delayed recall (as compared to undelayed recall) have been found using working memory delays as short as 900 milliseconds (Bae et al., 2015).

As mentioned above, more anterior regions might play a role in biasing working memory representations in sensory cortex to become more categorical over time. Our post-hoc analyses showed some evidence for working memory representations in frontal cortices (in left ventro-lateral prefrontal cortex and left premotor cortex). These regions have been shown in prior work to represent orientations (Edward F. Ester et al., 2015) and Chinese characters (Yan et al., 2021), respectively. In these regions, however, we found no significant differences between the variance explained by the two encoding models leaving it unclear what role these regions play in the representation of color during working memory.

Prior work has investigated categorical representations when subjects are explicitly instructed to categorize stimuli. When naming colors, for example, areas V1 and V4 demonstrate categorical clustering while no such clustering was evident when attention was diverted from the presented colors (Brouwer and Heeger, 2013). Similarly for orientation, motion and location stimuli, prior work (Ester et al., 2020; Freedman et al., 2001; Freedman and Assad, 2006) has shown that when subjects are trained to categorize stimuli into two arbitrary classes neural representations across visual, parietal, and prefrontal cortices can exhibit categorical properties. One fMRI study in particular, showed that orientation selective responses in early visual cortex are biased towards the center of the category (Ester et al., 2020). This finding resembles changes in orientation representations due to mental rotation (Albers et al., 2013) giving rise to the question whether categorical responses are a result of entrained categorical tuning or a rotation-like mental operation that maximizes the discriminability of the stimulus with respect to category (see discussion in Ester et al., 2020). Here, however, we demonstrate that mnemonic representations are categorical in visual cortex in the absence of any explicit or implicit instruction to categorize the colors and while subjects perform a task that actively encourages them to answer precisely. The similarity with categorical responses during color naming (Brouwer and Heeger, 2013), however, suggests that subjects are silently naming color stimuli during working memory as a rehearsal or encoding strategy. Prior work has suggested that such elaboration can improve performance in working memory tasks (Souza et al., 2021; Souza and Oberauer, 2018). Prior work has only found categorical biases in a non-categorization working memory task in prefrontal cortex where spatial responses appear to resemble the quadrant structure of the visual field (Leavitt et al., 2018).

In V1, we found that neither model outperformed the other. Prior work (Brouwer and Heeger, 2009) has suggested that V1 seems to harbor a fundamentally different discontinuous code (with respect to hue) based on opponent colors which is not reflected by either model. Further research is needed to investigate the vast space of potential color encoding models to give more insight into the properties and determinants of neural coding during working memory and perception.

Declaration of Competing Interest

The authors have no interests to declare.

Credit authorship contribution statement

Chang Yan: Conceptualization, Investigation, Data curation, Software, Formal analysis, Visualization, Writing – original draft. **Thomas B. Christophel:** Conceptualization, Software, Formal analysis, Visualization, Writing – original draft, Supervision. **Carsten Allefeld:** Software, Formal analysis, Writing – review & editing. **John-Dylan Haynes:** Conceptualization, Writing – review & editing, Supervision.

Data availability

All data analyzed in the current study including the analysis code are available to researchers from the corresponding author upon request. The data cannot be shared publicly, because subjects did not provide informed consent for public sharing. Sharing of data requires a formal data sharing agreement.

Acknowledgments

This work was funded by the Bernstein Computational Neuroscience Program of the German Federal Ministry of Education and Research BMBF Grant 01GQ0411, by the Excellence Initiative of the German Federal Ministry of Education and Research and DFG Grants GSC86/1-2009, KFO247, CH 1674/2-1, HA 5336/1-1, SFB 940 and JA 945/3-1/SL185/1-1.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2023.120149.

References

Albers, A.M., Kok, P., Toni, I., Dijkerman, H.C., de Lange, F.P., 2013. Shared representations for working memory and mental imagery in early visual cortex. Curr. Biol. 23, 1427–1431.

Allefeld, C., Haynes, J.-D., 2014. Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. Neuroimage 89, 345–357. doi:10.1016/j.neuroimage.2013.11.043.

Awh, E., Barton, B., Vogel, E.K., 2007. Visual working memory represents a fixed number of items regardless of complexity. Psychol. Sci. 18, 622–628.

Baddeley, A.D., 1986. Working Memory. Clarendon Press, Oxford (UK).

Bae, G.-Y., Olkkonen, M., Allred, S.R., 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. doi:10.1037/xge0000076.

Bays, P.M., Catalao, R.F.G., Husain, M., 2009. The precision of visual working memory is set by allocation of a shared resource. J. Vis. 9. doi:10.1167/9.10.7, 7.1-711.

Berlin, B., Kay, P., 1969. Basic Color Terms: Their Universality and Evolution. University of California Press, Berkeley and Los Angeles.

Bickel, P.J., Freedman, D.A., 1981. Some asymptotic theory for the bootstrap. Ann. Stat. 9, 1196–1217. doi:10.1214/aos/1176345637.

Brouwer, G.J., Heeger, D.J., 2009. Decoding and reconstructing color from responses in human visual cortex. J. Neurosci. 29, 13992–14003. doi:10.1523/JNEUROSCI.3577-09.2009.

Buschman, T.J., Siegel, M., Roy, J.E., Miller, E.K., 2011. Neural substrates of cognitive capacity limitations. Proc. Natl. Acad. Sci 108, 11252–11255. doi:10.1073/pnas.1104666108.

Christophel, Thomas B, Allefeld, C., Endisch, C., Haynes, J.-D., 2018a. View-independent working memory representations of artificial shapes in prefrontal and posterior regions of the human brain. Cereb. Cortex 28, 2146–2161. doi:10.1093/cercor/bhx119.

Christophel, T.B., Hebart, M.N., Haynes, J.-D., 2012. Decoding the contents of visual short-term memory from human visual and parietal cortex. J. Neurosci. 32, 12983–12989. doi:10.1523/JNEUROSCI.0184-12.2012.

Christophel, T.B., Iamshchinina, P., Yan, C., Allefeld, C., Haynes, J.-D., 2018b. Cortical specialization for attended versus unattended working memory. Nat. Neurosci. 21, 494–496.

Christophel, T.B., Klink, P.C., Spitzer, B., Roelfsema, P.R., Haynes, J.-D., 2017. The distributed nature of working memory. Trend. Cogn. Sci. 21, 111–124. doi:10.1016/j.tics.2016.12.007.

Cosgrove, K.P., Mazure, C.M., Staley, J.K., 2007. Evolving knowledge of sex differences in brain structure, function, and chemistry. Biol. Psychiatry Bipolar Disord. OCD: Circuit. Impul. Compul. Behav. 62, 847–855. doi:10.1016/j.biopsych.2007.03.001.

Dotson, N.M., Hoffman, S.J., Goodell, B., Gray, C.M., 2018. Feature-based visual short-term memory is widely distributed and hierarchically organized. Neuron 99, 215–226.e4. doi:10.1016/j.neuron.2018.05.026.

Efron, B., 1979. Bootstrap methods: another look at the Jackknife. Ann. Stat. 7, 1–26. doi:10.1214/aos/1176344552.

Ester, E.F., Anderson, D.E., Serences, J.T., Awh, E., 2013. A neural measure of precision in visual working memory. J. Cogn. Neurosci. 25, 754–761. doi:10.1162/jocn_a_00357.

Ester, E.F., Sprague, T.C., Serences, J.T., 2015. Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. Neuron 87, 893–905. doi:10.1016/j.neuron.2015.07.013.

Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2003. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J. Neurosci. 23, 5235–5246.

Freeman, W.T., Adelson, E.H., 1991. The design and use of steerable filters. IEEE Trans. Pattern Anal. Mach. Intell. 13, 891–906. doi:10.1109/34.93808.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S., 1994. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Fuster, J.M., 1997. Network memory. Trend. Neurosci. 20, 451–459.

Gladwin, T.E., 2020. An implementation of N-way repeated measures ANOVA: effect coding, automated unpacking of interactions, and randomization testing. MethodsX 7, 100947. doi:10.1016/j.mex.2020.100947.

Hernández, A., Nácher, V., Luna, R., Zainos, A., Lemus, L., Alvarez, M., Vázquez, Y., Camarillo, L., Romo, R., 2010. Decoding a perceptual decision process across cortex. Neuron 66, 300–314. doi:10.1016/j.neuron.2010.03.031.

Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2.

Kumar, S., Joseph, S., Gander, P.E., Barascud, N., Halpern, A.R., Griffiths, T.D., 2016. A brain system for auditory working memory. J. Neurosci. 36, 4492–4505. doi:10.1523/JNEUROSCI.4341-14.2016.

Leavitt, M.L., Pieper, F., Sachs, A.J., Martinez-Trujillo, J.C., 2018. A quadrantic bias in prefrontal representation of visual-mnemonic space. Cereb. Cortex N. Y. N 1991 28, 2405–2421. doi:10.1093/cercor/bhx142.

Lee, S.-H., Kravitz, D.J., Baker, C.I., 2013. Goal-dependent dissociation of visual and prefrontal cortices during working memory. Nat. Neurosci. 16, 997–999. doi:10.1038/nn.3452.

Lindsey, D.T., Brown, A.M., 2006. Universality of color names. Proc. Natl. Acad. Sci. U. S. A. 103, 16608–16613. doi:10.1073/pnas.0607708103.

Loreto, V., Mukherjee, A., Tria, F., 2012. On the origin of the hierarchy of color names. Proc. Natl. Acad. Sci. 109, 6819–6824. doi:10.1073/pnas.1113347109.

Luck, S.J., Vogel, E.K., 1997. The capacity of visual working memory for features and conjunctions. Nature 390, 279–280.

Panichello, M.F., DePasquale, B., Pillow, J.W., Buschman, T.J., 2019. Error-correcting dynamics in visual working memory. Nat. Commun. 10, 3366. doi:10.1038/s41467-019-11298-3.

Rademaker, R.L., Chunharas, C., Serences, J.T., 2019. Coexisting representations of sensory and mnemonic information in human visual cortex. Nat. Neurosci. 22, 1336–1344. doi:10.1038/s41593-019-0428-x.

Salazar, R.F., Dotson, N.M., Bressler, S.L., Gray, C.M., 2012. Content-specific Fronto-parietal synchronization during visual working memory. Science 338, 1097–1100. doi:10.1126/science.1224000.

Singh, K., 1981. On the asymptotic accuracy of Efron's bootstrap. Ann. Stat. 9, 1187–1195. doi:10.1214/aos/1176345636.

Siuda-Krzywicka, K., Witzel, C., Chabani, E., Taga, M., Coste, C., Cools, N., Ferrieux, S., Cohen, L., Seidel Malkinson, T., Bartolomeo, P., 2019. Color categorization independent of color naming. Cell Rep. 28, 2471–2479.e5. doi:10.1016/j.celrep.2019.08.003.

Sperling, G., 1960. The information available in brief visual presentations. Psychol. Monogr. Gen. Appl. 74, 1–29.

Sprague, T.C., Ester, E.F., Serences, J.T., 2014. Reconstructions of information in visual spatial working memory degrade with memory load. Curr. Biol. 24, 2174–2180. doi:10.1016/j.cub.2014.07.066.

Sprague, T.C., Serences, J.T., 2013. Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. Nat. Neurosci. 16, 1879–1887. doi:10.1038/nn.3574.

Wang, L., Mruczek, R.E.B., Arcaro, M.J., Kastner, S., 2015. Probabilistic maps of visual topography in human cortex. Cereb. Cortex 25, 3911–3931. doi:10.1093/cercor/bhu277.

Wilken, P., Ma, W.J., 2004. A detection theory account of change detection. J. Vis. 4, 11. doi:10.1167/4.12.11, 11.

Yan, C., Christophel, T.B., Allefeld, C., Haynes, J.-D., 2021. Decoding verbal working memory representations of Chinese characters from Broca's area. Neuroimage 226, 117595. doi:10.1016/j.neuroimage.2020.117595.