Decoding visual colour from scalp electroencephalography measurements

Jasper E. Hajonides[1,2,@], Anna C. Nobre[1,2], Freek van Ede[1,3]*, Mark G. Stokes[2]*

[1]*Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, UK.*

[2]*Department of Experimental Psychology, University of Oxford, Oxford, UK.*

[3]*Institute for Brain and Behavior Amsterdam, Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, The Netherlands*

Abstract

Recent advances have made it possible to decode various aspects of visually presented stimuli from patterns of scalp EEG measurements. As of recently, such multivariate methods have been commonly used to decode visual-spatial features such as location, orientation, or spatial frequency. In the current proof-of-principle study, we show that it is also possible to track visual colour processing by using Linear Discriminant Analysis. Building on other recent demonstrations, we show that colour decoding: (1) reflects sensory qualities (as opposed to, for example, verbal labelling) with a prominent contribution from posterior electrodes contralateral to the stimulus, (2) conforms to a parametric coding space, (3) is possible in multi-item displays, and (4) is comparable in magnitude to the decoding of visual stimulus orientation. Through subsampling our data, we also provide an estimate of the approximate number of trials and participants required for robust decoding. Colour decoding opens a relevant new dimension in which to track visual processing using scalp EEG measurements, while bypassing potential confounds associated with decoding approaches that focus on spatial features.

* these authors contributed equally.

@ Corresponding author at: *Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, UK*

Email: *jasper.hajonidesvandermeulen@new.ox.ac.uk*

Introduction

Human scalp electroencephalography (EEG) is sensitive to synchronous activity in large neural populations, thus providing a macroscopic readout of brain activity. Whereas extracranial measures of brain activity carry the problem of source ambiguity (Koles, 1998), it is nevertheless possible to compare scalp topography of brain activity to differentiate spatial patterns generated by subtly different configurations of neuronal populations. Recently, studies have shown that the information across an array of electrodes can be used in multivariate analyses to decode neural information such as stimulus orientation, spatial location, spatial frequency, and motion direction (e.g. Ramkumar *et al.*, 2013; King and Dehaene, 2014; Cichy, Ramirez and Pantazis, 2015; Myers *et al.*, 2015; Stokes, Wolff and Spaak, 2015; Foster *et al.*, 2017; Bae and Luck, 2019; Sandhaeger *et al.*, 2019).

The use of scalp EEG for tracking colour processing remains less investigated and applied. In contrast to the type of visual-spatial features that have typically been the focus in EEG decoding studies to date (such as orientation, location, and spatial frequency), the visual feature *colour* is not defined by spatial parameters. Accordingly, it should be less susceptible to contributions of spatial attention and/or eye movements, such as spatial biases in micro-saccades (Hafed and Clark, 2002; Engbert and Kliegl, 2003; Hollingworth, Matsukura and Luck, 2013; Mostert *et al.*, 2018; Quax *et al.*, 2019; Thielen *et al.*, 2019; van Ede, Chekroud and Nobre, 2019). The demonstration of colour decoding from scalp EEG should therefore have important implications – affording novel ways to track perceptual representations in various contexts.

Like orientation, colour also follows an organisational structure in visual cortices (Kleinschmidt *et al.*, 1996; Engel, Zhang and Wandell, 1997), and research has shown that it is possible to decode colour with functional MRI measurement with millimetre resolution (Brouwer and Heeger, 2009) across a range of cortical visual areas (Brouwer and Heeger, 2009; Persichetti *et al.*, 2015). Complementing these findings, a handful of recent initial studies have suggested the possibility of decoding visual colour information from scalp EEG measurements (Bocincova and Johnson, 2018; Sandhaeger *et al.*, 2019) and related MEG measurements (e.g. Sandhaeger *et al.*, 2019; Hermann *et al.*, 2020).

To further establish the feasibility, and characterise the properties, of colour decoding from scalp EEG measurements, it remains to be determined that the relevant signals represent the sensory processing of colours and are independent from other associated factors, such as verbal labels or decision-making in tasks. To this end, it would prove reassuring and advantageous if decoding of colour originated from posterior sites contralateral to the decoded stimulus, and followed parametric variations in stimuli (as opposed to binary or categorical boundaries) in a manner akin to decoding of stimulus orientation. In addition, the ability to decode the colours of multiple stimuli presented concurrently (as opposed to a single central stimulus in isolation, as in prior colour decoding studies) would greatly add to the flexibility and analytical power of colour decoding as a tool in future studies.

In the present proof-of-principle study, we therefore assessed the feasibility of tracking parametric variations in colour processing in scalp EEG, and qualitatively compared 'colour decoding' to the more established decoding of stimulus orientation (in term of robustness, tuning profile, timing, spatial topography, and so on). To do so, we presented 30 participants with two sinusoidal Gabor gratings (one left, one right) that each contained a unique colour and orientation, and we required participants to attend to both features of both items (for a later memory test, which is beyond the scope of the current study). Using Linear Discriminant Analysis, we obtained reliable (and comparable) decoding of the colour and orientation of both stimuli. Our data also provide a clear

case for a "visual" account of this decoding, by showing that colour decoding has a contralateral-posterior topography, and follows a parametric colour-coding space.

Materials & Methods

Participants

Ethical approval for this study was obtained from the Central University Research Ethics Committee of the University of Oxford. Thirty healthy adults (28.3 years old on average; ranging from 18-35 years old; 17 females) were included in this study after excluding two participants due to technical difficulties and incomplete data collection. All participants had normal or corrected-to-normal vision and reported no deficiencies in their colour vision. All participants were reimbursed £15 per hour for their time.

Visual display

Stimuli were generated using the Psychophysics Toolbox version 3.0.11 (Brainard, 1997) in MATLAB 2014b (MathWorks, Natick, MA). The stimuli were presented on a 23-inch monitor (1920 × 1080, 100Hz). Viewing distance was 90 cm from the monitor where participants were stabilised at a fixed position using a chin rest.

Stimuli & design

The critical stimuli for this experiment were presented within the context of a visual working-memory task that ensured attentive viewing of the visual stimuli at encoding. Task details are presented here for clarity, but the analyses only focus on decoding the memory stimuli at encoding (for a full description of the task and participants' behavioural performance, see: Hajonides *et al.*, 2020). At the start of each trial, two Gabor stimuli with a diameter of 4.3 degrees of visual angle (one to the left, the other to the right of fixation, centred at ±4.3 degrees each) were presented simultaneously for 300 ms (**Figure 1A**) on a grey background (RGB: 0.196 0.196 0.196). The stimuli consisted of bilateral luminance-defined sinusoidal Gabor gratings and each had a unique colour and orientation. Colours were drawn from a CIELAB colour space, which is often used to define perceptually uniform stimulus sets. Forty-eight evenly-spaced colours were drawn from a circle in CIELAB colour space with a fixed lightness of L = 54 (**Figure 1B**; centre is at a = 18, b = -8, radius =59).

In total, 48 different colours and 48 different orientations were presented, with feature values ranging from 3.75 to 180 degrees in evenly spaced steps. Orientation and colour values varied independently of one another so that no two of the same colours or orientations were presented simultaneously. To increase the amount of samples for feature, we binned the 48 colour and orientation features into 12 - equally spaced – bins (**Figure 1B**).

At the time of stimulus presentation, participants were instructed to attend to and remember both the colour and the orientation of both gratings. At the end of each trial, participants were probed to report a single feature from one of the two items. Which feature and item would be probed was unpredictable at encoding. During the 2400-ms memory delay, two events occurred: (1) 1000 ms after encoding onset a cue could provide information about the item or feature dimension that would be probed and (2) 1800 ms after encoding onset, a colour or orientation distractor was presenting to distort visual processing. These trial details are irrelevant to the current investigation, which focuses exclusively on the encoding phase of the experiment.

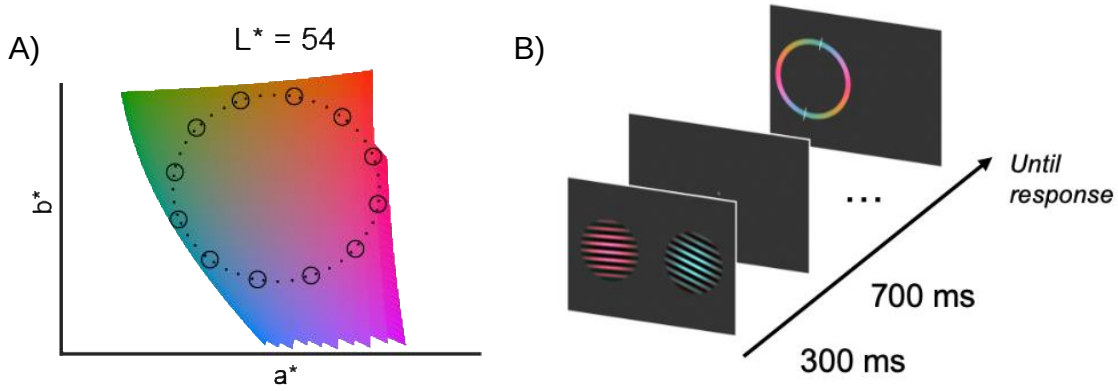


Figure 1. Stimulus characteristics and presentation. A) Forty-eight colours - depicted in small dots - were sampled from CIELAB colour space with a fixed lightness ("L") of 54. During analysis, colours were binned into 12 evenly-spaced bins, depicted in circles. B) Participants were presented with two Gabor gratings for 300 ms with both colour and orientation information. Colour and orientation were varied independently, using circular spaces with 48 possible feature-values (indicated in panel A). To ensure attentive encoding of the stimuli, participants were required to reproduce the exact colour or orientation of either stimulus after a brief memory delay (the schematic shows an example trial prompting a report for the colour of the left grating). We focus our analyses on the 1000 ms after onset of the gratings. See methods for the full details of the task design.

EEG recording

The EEG signal was acquired using a 10-20 system using 61 Ag/AgCl sintered electrodes (EasyCap, Herrsching, Germany). The data were recorded using a Synamps amplifier and Curry 8 acquisition software (Compumedics NeuroScan, Charlotte, NC). An electrode placed behind the right mastoid was used as the active reference during acquisition. Offline, data were re-referenced to the average of the left and right mastoids. The ground was placed above the left elbow. Bipolar electrooculography (EOG) was recorded from electrodes above and below the left eye and lateral of both eyes. EEG was collected at 1000 Hz and impedances were kept below 7kΩ.

EEG pre-processing

EEG data were analysed in MATLAB 2017a using FieldTrip (Oostenveld *et al.*, 2011) in conjunction with the OHBA Software Library (OSL; https://ohba-analysis.github.io/). After importing, the data were epoched from -300 ms to 1400 ms around presentation the presentation of the Gabor gratings (ft_redefinetrial) and re-referenced to the average of both mastoids (ft_preprocessing). Raw epoched data were downsampled to 200 Hz and stored for subsequent analyses (analysis scripts and data will be made available online at OSF https://osf.io/j289e/?view_only=b13407009b4245f7950960c34a5474a6).

Subsequently, we used the EOG to identify trials in which participants blinked and might have missed the visually presented stimuli. The vertical EOG was baselined between 300 ms to 100 ms before stimulus presentation. Trials that exceeded half of the maximum voltage elicited by blinks (half of 400 μV) within the timeframe of 100 ms prior to and 100 ms post stimulus presentation were marked and later discarded. Additional eye-related artefacts were identified using Independent Component Analysis (ICA; ft_componentalanalysis) by applying the FastICA algorithm (Hyvärinen, 1999) to the full EEG layout. Components were rejected if they had a correlation of $r > 0.40$ with the EOG electrodes (on average 2.7 components per participant, and a maximum of 4 components). In addition to trials where participants blinked at the moment of stimulus presentation, trials with high within-trial variance of the EEG broadband signal were

removed using a generalised ESD test at a 0.05 significance threshold (Rosner, 1983; implemented in OSL), removing 2.48 % ± 2.18 % (mean ± standard deviation) of all trials.

Classification on EEG data

To focus our decoding on visual processing, we restricted our main analyses on the 17 most posterior electrodes (P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz and O2; as in Wolff *et al.*, 2017).

To make class predictions we employed Linear Discriminant Analysis (LDA) using the Scikit-learn toolbox (Pedregosa *et al.*, 2011) and estimated the likelihood for all classes. LDA is a multi-class decoding algorithm that performs well for neuroscientific time-course data (Grootwagers, Wardle and Carlson, 2017).

To increase the sensitivity of our decoding analysis, we capitalised on both spatial and temporal patterns for decoding (Grootwagers, Wardle and Carlson, 2017; Wolff *et al.*, 2020). To do so, we used a sliding-window approach and concatenated topographical patterns from t_0 up to t_{-19} (20 steps of 10 ms, ranging over 200 ms) into a single vector. This increased the amount of features in the LDA classifier from 17 (17 posterior electrodes) to 340 (17 electrodes × 20 time steps). To discount absolute amplitude differences between trials, we demeaned information within each 200-ms sliding window. This analysis method thus exploits embedded informative temporal variability, reduces noise by combining data over time, and circumvents baselining issues (Grootwagers, Wardle and Carlson, 2017; Wolff *et al.*, 2020). In a supplementary analysis, we confirmed that similar (be it less robust) colour and orientation decoding could be obtained when decoding based on purely spatial patterns, on a timepoint-by-timepoint basis.

Principal component analysis was applied to the data to capture 95% of the variance in the data (sklearn.decomposition.PCA). Trial feature labels were binned into 12 classes (each subtending $12/\pi$; numpy.digitize). These data were split into train-and-test sets using 10-fold stratified cross-validation (sklearn.model_selection.RepeatedStratifiedKFold). Next, training data were standardised (sklearn.preprocessing.StandardScaler) and after the same standardisation was applied to the test data we fit the LDA to the training data and labels (sklearn.discriminant_analysis.LinearDisciminantAnalysis; using singular value decomposition and a default threshold for rank estimation of 0.001). Likelihoods for each class were estimated after training and testing the LDA classifier separately for each time point, features (colour and orientation from left and right sides), and participant. The resulting likelihoods were centred so that the predicted class was always in centre surrounded by neighbouring feature classes (…, $-\frac{1}{12}\times\pi$, 0, $+\frac{1}{12}\times\pi$, …). These predicted-class likelihoods were convolved with a cosine ($-\frac{1}{2}\pi$, $\frac{1}{2}\pi$) to provide a single measure of evidence. Training and testing were performed separately for left and right features and only at the final stage we averaged across both sides.

To confirm the posterior locus of decodability, we also ran a searchlight decoding analysis (Kriegeskorte, Goebel and Bandettini, 2006), iteratively considering all electrodes. To this end, we applied the same classification on the average voltage between 150 and 350 ms, but this time always only considered the data from a given electrode together with the data from its immediately adjacent (left/right/above/below) neighbours.

Statistical analysis

For the analysis of the classifier predictions, cosine-convolved evidence for each decoded feature-dimension (colour, orientation) was compared to data with labels randomly shuffled for each time

point using a cluster permutation test (mne.stats.permutation_cluster_test) (Gramfort *et al.*, 2014) with 10000 iterations and a F-statistic threshold of 2.045 ($p < .05$ with 29 degrees of freedom).

For further characterisation of the data, we also ran several analyses with data collapsed over the 150-350 post-stimulus time-window in which both colour and orientation decoding were particularly prominent. We first used this window to assess the reliability of the LDA likelihoods for each of the 12 colours and orientations separately. For each bin we tested decoding against zero using a one-tailed one-sample t-tests, and report both the uncorrected results and the results after applying a Bonferroni correction for multiple comparisons. We also used the data from the same window to investigate the parametric coding spaces for colour and orientation by applying multi-dimensional scaling (MDS; sklearn.manifold.MDS; Kruskal, 1964) to a matrix with rows for decodable colour or orientation and columns for classifier predictions. This matrix was inverted so that the diagonal had the smallest likelihoods and we normalised each row by subtracting the lowest value of the row. We ran MDS with 1000 initialisations and a maximum of 1000 iterations to find the configuration with the lowest stress score for a two-dimensional embedding of the 12 colours and orientations. Finally, we varied the number of participants and trials/participant we included in our analysis. Doing so, we randomly sampled a subset of participants (5-30) and the first *n* trials/participants (where *n* ranged from 96 to 912 in steps of 48), we took this approach to simulate having run a shorter experiment. Subsequently, we evaluated the cosine-convolved classifier evidence between 150 ms and 350 ms for colours and orientations, averaged over left and right stimulus presentations. To ensure random sampling we drew each subset of participants over 100 permutations and averaged over the results. In contrast to other analyses, a 5-fold cross validation was used to deal with the full range of trial numbers. The resulting matrix was slightly smoothed using a gaussian kernel with a sigma of 0.5 (scipy.ndimage.gaussian_filter), where edge artefacts were avoided by replicating the nearest data point.

Results

Colour can be decoded from scalp EEG, and is comparable to orientation decoding

We extracted EEG data from the 17 posterior electrodes ranging from 200 ms prior to the presentation of the two gratings to 1000 ms after stimulus presentation onset. Average ERP traces revealed highly similar voltage profiles across the range of tested colours (**Figure 2A**, left) and orientations (**Figure 2A**, right). No systematic or significant differences occurred in the ERPs as a function of colour or orientation. Yet, an LDA classifier was able to uncover a clear reconstruction of the presented colours and as well orientations (**Figure 2B**; averaged over left and right stimulus presentations). To increase sensitivity and to ease visualisation, we reduced these tuning-profiles from **Figure 2B,** to a single decodability value per time point, by convolving the tuning profiles with a cosine function (as in Wolff *et al.*, 2017; van Ede *et al.*, 2018; Wolff *et al.*, 2019). Decoding time-courses are depicted in **Figure 2C**. We found highly significant decodability for both visual features. Colour decoding first reached significance after 115 ms and remained significant until 875 ms. Orientation decoding first reached significance after 135 ms and remained significant until 945 ms.

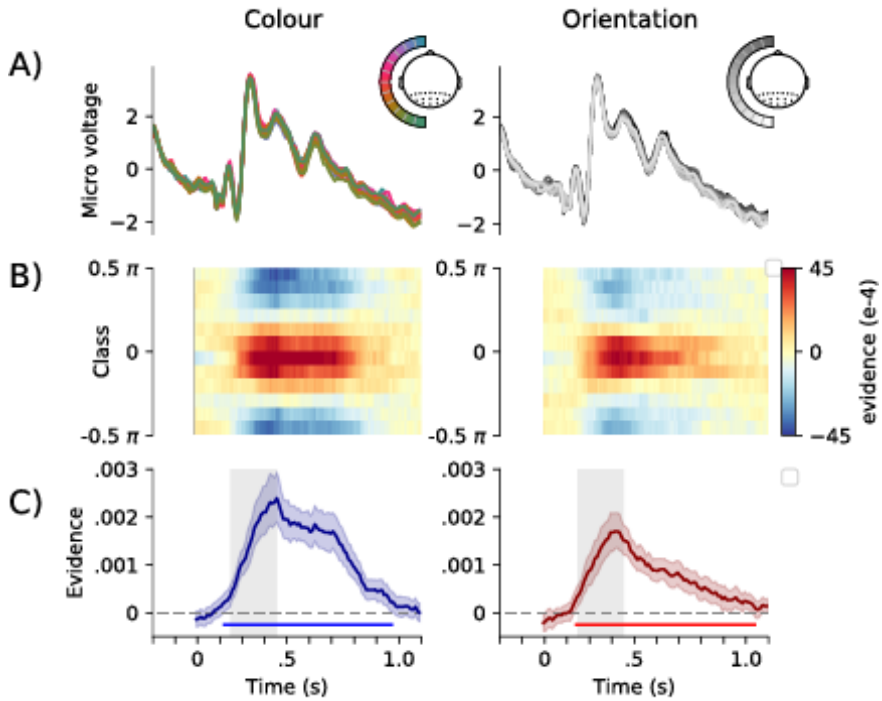


Figure 2. ERPs and decoding performance for colours and orientations. A) Event-related potentials for the posterior 17 electrodes (see inset), locked to stimulus presentation and displaying all twelve different orientations or colours. B) Two-dimensional, time-resolved, tuning curves with the likelihood for target colour or orientation (Class "0") and all neighbouring values of theta. C) Mean cosine-convolved evidence with red lines for colour, blue for orientation. Cluster-permutation corrected significant time points are indicated with horizontal lines. Grey-shaded area between 150 to 350 ms is used for subsequent analyses. Error bars show 95% confidence intervals, calculated across participants (n = 30).

Colour decoding is driven by visual signals with a contralateral-posterior topography

To understand what contributed to our ability to decode colour, we next ran a searchlight decoding analysis (Kriegeskorte, Goebel and Bandettini, 2006) across all sensors (van Ede *et al.*, 2019; Wolff *et al.*, 2020). This demonstrated that classifier evidence (in the 150 to 350 post-stimulus period in which we found robust decoding; grey-shaded windows in **Figure 2C**) peaked in posterior

electrodes, with primary contributions from electrodes contralateral to the decoded stimuli. This was the case both for decoded colour (**Figure 3A**), and orientation (**Figure 3B**). This provides compelling evidence for the "visual" nature of this decoding (as opposed to, for example, decoding distinct "verbal labels" associated with distinct colours).

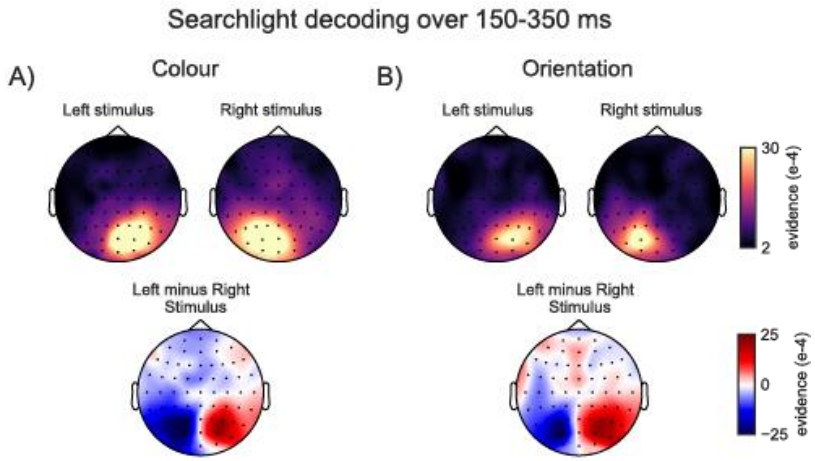


Figure 3. Colour and orientation decoding are primarily driven by posterior electrodes contralateral to the decoded stimulus. Top row topographies show which electrodes show highest mean cosine-convolved evidence for A) colours and B) orientations, of stimuli that were presented on the left or on the right of fixation. Posterior contralateral electrodes show highest evidence. Bottom panels show the difference between left-stimulus and right-stimulus decoding topographies, highlighting the lateralisation of the evidence, and the comparable lateralisation obtained for colour and orientation decoding.

Colour decoding is parametrically organised

To test how well each of the feature-value bins could be decoded (relative to all other feature-value bins within the colour and orientation spaces), we analysed mean cosine-convolved LDA likelihoods in the data pooled between 150-ms and 350-ms window that showed robust decoding of both features (indicated in grey in **Figure 2C**). Decoding, as a function of the decoded feature-values, is shown in **Figure 4A**. For colour bins, all 12/12 colours could be decoded significantly against a $p < .05$ threshold (mean $t_{29} = 4.523$; min = 2.440; max = 6.175), and for orientation 11/12 orientations could be decoded significantly against an uncorrected single-tailed threshold (mean $t_{29} = 4.096$; min = 1.585; max = 7.390). After applying Bonferroni correction, 9/12 colour bins were significant and 8/12 orientations. Using multi-dimensional scaling we revealed the representational space in two dimensions, showing a near-circular configuration for both colour and orientation bins (**Figure 4B**). For orientation, the cardinal orientations (0, 90, and 180 degrees) showed higher decoding than the off-cardinal decoding.

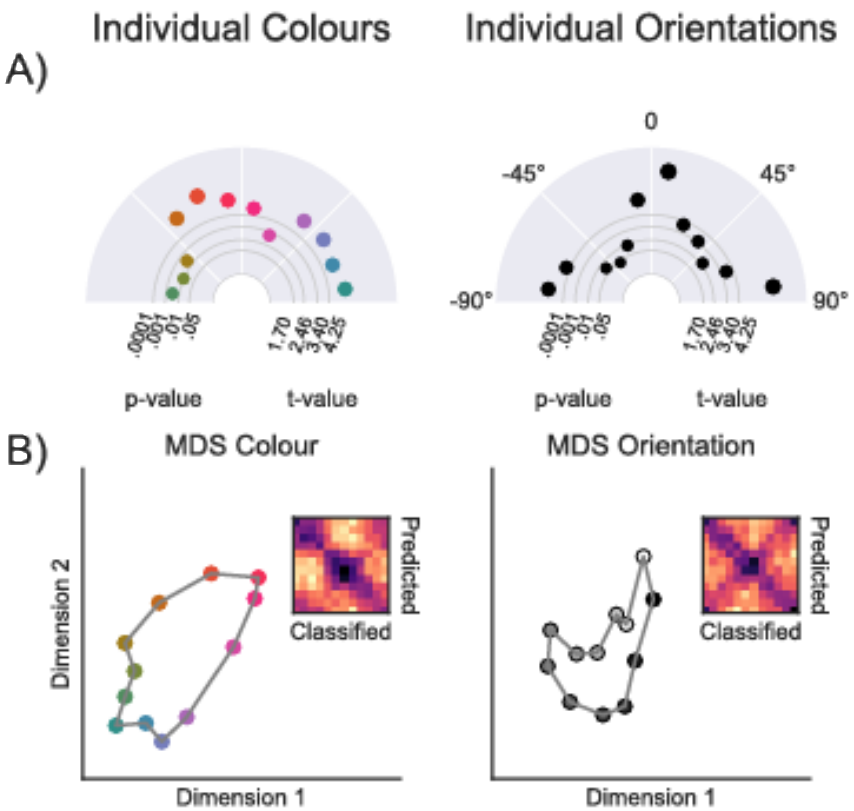


Figure 4. Decoding of individual features between 150 and 350 ms. A) This polar plot illustrates the decoding evidence of 12 individual colours and orientations. The distance from the centre depicts the t-value of testing the cosine-convolved tuning curve for a single feature, left and right combined, across participants, relative to zero. The size of the dots mirrors the magnitude of the evidence for that feature B) Two-dimensional visualisation of the similarity matrix in which we observed a circular configuration for colour features. Dissimilarity matrices show feature (colour or orientation) on the y-axis and distance between bins of the tuning curve on the x-axis.

Dependence of decoding on trial and participant numbers

The current study included 30 participants who each completed over 900 trials. To provide guidance on "how many" participants and trials would typically be required to achieve robust decoding, we ran the following analysis. We randomly sampled *x* participants (in 100 permutations) and, for each included participant, we incrementally increased the number of trials used for our decoding analysis, always including the first *y* trials of the session (as if we had only collected data up to this trial). For each combination of # participant by # trials per participant, we assessed feature decoding between 150 to 350 ms after stimulus onset, and plotted the resulting matrix (**Figure 5**). Both the number of trials and the number of participants have a strong effect on the reliability of feature decoding. It is evident that both sources of statistical power influence our decoding scores.

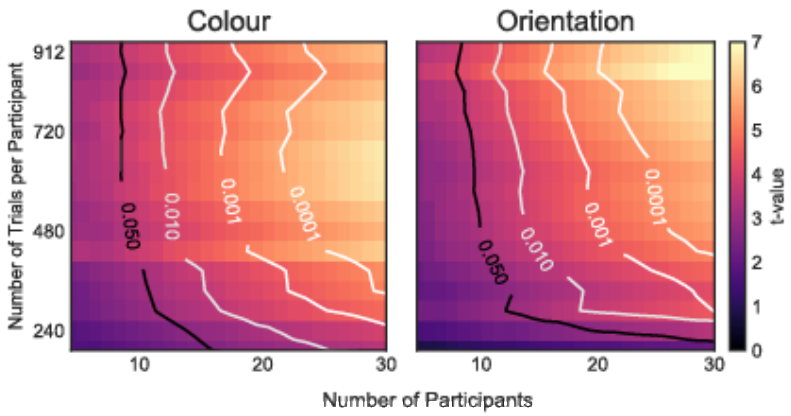


Figure 5. Feature decoding as a function of trial and participant numbers. Average cosine-convolved evidence between 150 and 350 ms for the average of left and right feature decoding. Contours mark the .05, .01, .001, and .0001 p-value thresholds for n-1 degrees of freedom.

Discussion

We investigated the ability to decode visual colours from scalp EEG measurements. We built on other recent studies that have employed colour decoding in scalp EEG (Bocincova and Johnson, 2018; Sandhaeger *et al.*, 2019) and MEG (Sandhaeger *et al.*, 2019; Hermann *et al.*, 2020) and extended these in several ways. We show that we can reliably classify four simultaneously presented visual features – two colours and two orientations – from posterior EEG recordings using multi-class LDA. Follow-up analyses revealed that activity in posterior electrodes contralateral to the decoded stimulus were the primary contributors to the decoding of both features, suggesting that visual sensory processing was the main source of decodable signals, ruling out alternative explanations of colour decoding, such as verbal labelling. Furthermore, colour decoding followed a colour circle in representational space, speaking to the parametric nature of visual colour decoding. Finally, by having both orientation and colour information present in our stimuli, we were able to decode both. This revealed how the decoding of colour is comparable in magnitude, and if anything even clearer, to the decoding of orientation in a visual stimulus.

To increase sensitivity, in the current work we capitalised on both spatial and temporal activity patterns for classification (as in Grootwagers, Wardle and Carlson, 2017; Wolff *et al.*, 2020). However, to ensure that our results did not critically depend on this choice, we also ran all of our analysis using a more basic decoding approach whereby we relied only on spatial patterns for decoding; running the decoding analysis separately for each time point after stimulus onset. While this analysis was, as expected, less sensitive, we could nevertheless demonstrate that it was still possible to significantly decode visual colour from purely spatial patterns in the EEG data; that colour and orientation decoding remained highly comparable in magnitude; and that colour and orientation decoding each had a contralateral-posterior topography and conformed to a parametric coding space (**Supplementary Figures 1-4**).

Perceived colour and luminance depend on many factors, including monitor setup, local stimulus contrasts, and individual biology (Wachtler and Wehrhahn, 1996). While it is not straightforward to disentangle the respective contributions of each of these factors to our ability to decode colour in the current study, we believe our colour decoding is unlikely due to possible low-level differences in luminance between the different colours that we used. First, stimulus colours were sampled from a section of CIELAB space with equal lightness. Second, even if luminance differences contributed, they would be unlikely to result in the parametric coding space defined

for colour that we observed. Third, we found no differences in ERP morphology or magnitude between colours, suggesting all colours gave rise to similar univariate responses. Moreover, other related research has recently described how luminance and hue may modulate brain activity independently (Hermann *et al.*, 2020).

The ability to decode colour, as a reliable proxy of detailed visual processing, is important not only because it opens a new dimension through which to track detailed visual processing on the basis of scalp EEG measurements. It may also prove instrumental to by-pass a fundamental limitation faced by other decodable features such as spatial location, orientation, or spatial frequency. Colour, being a non-spatial parametric feature, should be much less susceptible to confounds associated with visual-spatial processing, such micro-saccades (Hafed and Clark, 2002; Engbert and Kliegl, 2003; Hollingworth, Matsukura and Luck, 2013; Mostert *et al.*, 2018; Quax *et al.*, 2019; Thielen *et al.*, 2019; van Ede, Chekroud and Nobre, 2019).

In conclusion, like several other recent studies, we show that colour decoding is possible from scalp EEG measurements. Building on this related recent work, we have shown that this colour decoding reflects visual processing with a clear posterior-contralateral topography; that it conforms to a parametric colour-coding space; that it is possible in multi-item display; and that it is comparable to the decoding of stimulus orientation. This opens a relevant new dimension in which to track visual processing using scalp EEG measurements, while bypassing potential confounds associated with conventional decoding approaches that focus on spatial features.

Acknowledgements

This research was funded by an ESRC Grand Union studentship and the Scatcherd European Scholarship awarded to J.E.H., a Marie Skłodowska-Curie Fellowship from the European Commission (ACCESS2WM) to F.v.E., and was supported by a James S. McDonnell Foundation Scholar Award (220020405), an ESRC grant (ES/S015477/1), the Medical Research Council Career Development Award (MR/J009024/1), and a Biotechnology and Biological Sciences Research Council award (BB/M010732/1) to M.G.S., as well as a James S. McDonnell Foundation Understanding Human Cognition Collaborative Award (number 220020448), and a Wellcome Trust Senior Investigator Award (104571/Z/14/Z) to A.C.N., and was supported by the NIHR Oxford Health Biomedical Research Centre. The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203139/Z/16/Z). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The authors declare that no competing interests exist.

References

Bae, G. Y. and Luck, S. J. (2019) 'Decoding motion direction using the topography of sustained ERPs and alpha oscillations', *NeuroImage*. Elsevier Ltd, 184(August 2018), pp. 242–255. doi: 10.1016/j.neuroimage.2018.09.029.

Bocincova, A. and Johnson, J. S. (2018) 'The time course of encoding and maintenance of task-relevant versus irrelevant object features in working memory', *Cortex*. doi: 10.1016/j.cortex.2018.10.013.

Brainard, D. H. (1997) 'The Psychophysics Toolbox', *Spatial Vision*, 10, pp. 433–436. doi: 10.1163/156856897X00357.

Brouwer, G. J. and Heeger, D. J. (2009) 'Decoding and reconstructing color from responses in human visual cortex', *Journal of Neuroscience*, 29(44), pp. 13992–14003. doi: 10.1523/JNEUROSCI.3577-09.2009.

Cichy, R. M., Ramirez, F. M. and Pantazis, D. (2015) 'Can visual information encoded in cortical columns be decoded from magnetoencephalography data in humans?', *NeuroImage*. Elsevier Inc., 121, pp. 193–204. doi: 10.1016/j.neuroimage.2015.07.011.

Engbert, R. and Kliegl, R. (2003) 'Microsaccades uncover the orientation of covert attention', *Vision Research*, 43(9), pp. 1035–1045. doi: 10.1016/S0042-6989(03)00084-1.

Engel, S., Zhang, X. and Wandell, B. (1997) 'Colour tuning in human visual cortex measured with functional magnetic resonance imaging', *Nature*, 388(6637), pp. 68–71. doi: 10.1038/40398.

Foster, J. J. *et al.* (2017) 'Alpha-Band Activity Reveals Spontaneous Representations of Spatial Position in Visual Working Memory', *Current Biology*. Elsevier Ltd., 27(20), pp. 3216-3223.e6. doi: 10.1016/j.cub.2017.09.031.

Gramfort, A. *et al.* (2014) 'MNE software for processing MEG and EEG data', *NeuroImage*. Elsevier Inc., 86, pp. 446–460. doi: 10.1016/j.neuroimage.2013.10.027.

Grootwagers, T., Wardle, S. G. and Carlson, Th. A. (2017) 'Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data', *Journal of Cognitive Neuroscience*, 29(4), pp. 677–697. doi: 10.1162/jocn_a_01068.

Hafed, Z. M. and Clark, J. J. (2002) 'Microsaccades as an overt measure of covert attention shifts', *Vision Research*, 42(22), pp. 2533–2545. doi: 10.1016/S0042-6989(02)00263-8.

Hajonides, J. E. *et al.* (accepted) 'Comparing the prioritisation of items and feature-dimensions in visual working memory', *Journal of Vision*.

Hermann, K. *et al.* (2020) 'Temporal dynamics of the neural mechanisms for encoding hue and luminance contrast uncovered by magnetoencephalography', *BioRxiv*, pp. 1–37.

Hollingworth, A., Matsukura, M. and Luck, S. J. (2013) 'Visual Working Memory Modulates

Rapid Eye Movements to Simple Onset Targets', *Psychological Science*, 24(5), pp. 790–796. doi: 10.1177/0956797612459767.

Hyvärinen, A. (1999) 'Fast and robust fixed-point algorithms for independent component analysis', *IEEE Transactions on Neural Networks*, 10(3), pp. 626–634. doi: 10.1109/72.761722.

King, J. R. and Dehaene, S. (2014) 'Characterizing the dynamics of mental representations: The temporal generalization method', *Trends in Cognitive Sciences*. Elsevier Ltd, 18(4), pp. 203–210. doi: 10.1016/j.tics.2014.01.002.

Kleinschmidt, A. *et al.* (1996) 'Functional mapping of color processing by magnetic resonance imaging of responses to selective P- and M-pathway stimulation', *Experimental Brain Research*, 110(2), pp. 279–288. doi: 10.1007/BF00228558.

Koles, Z. J. (1998) 'Trends in EEG source localization', *Electroencephalography and Clinical Neurophysiology*, 106(2), pp. 127–137. doi: 10.1016/S0013-4694(97)00115-6.

Kriegeskorte, N., Goebel, R. and Bandettini, P. (2006) 'Information-based functional brain mapping', *Proceedings of the National Academy of Sciences of the United States of America*, 103(10), pp. 3863–3868. doi: 10.1073/pnas.0600244103.

Kruskal, J. B. (1964) 'Nonmetric multidimensional scaling: A numerical method', *Psychometrika*, 29(2), pp. 115–129. doi: 10.1007/BF02289694.

Mostert, P. *et al.* (2018) 'Eye movement-related confounds in neural decoding of visual working memory representations', *eNeuro*, 5(4), pp. 1–14. doi: 10.1523/ENEURO.0401-17.2018.

Myers, N. E. *et al.* (2015) 'Testing sensory evidence against mnemonic templates', *eLife*, 4(DECEMBER2015), pp. 1–25. doi: 10.7554/eLife.09000.

Oostenveld, R. *et al.* (2011) 'FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data', *Computational Intelligence and Neuroscience*, 2011. doi: 10.1155/2011/156869.

Pedregosa, F. *et al.* (2011) 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research*, 12, pp. 2825–2830. doi: 10.1007/s13398-014-0173-7.2.

Persichetti, A. S. *et al.* (2015) 'Functional magnetic resonance imaging adaptation reveals a noncategorical representation of hue in early visual cortex', *Journal of Vision*, 15(6), pp. 1–19. doi: 10.1167/15.6.18.

Quax, S. C. *et al.* (2019) 'Eye movements explain decodability during perception and cued attention in MEG', *NeuroImage*. Elsevier Ltd, 195(March), pp. 444–453. doi: 10.1016/j.neuroimage.2019.03.069.

Ramkumar, P. *et al.* (2013) 'Feature-specific information processing precedes concerted activation in human visual cortex', *Journal of Neuroscience*, 33(18), pp. 7691–7699. doi: 10.1523/JNEUROSCI.3905-12.2013.

Rosner, B. (1983) 'Percentage Outlier Points for Generalized ESD Many- Procedure', *Technometrics,* 25(2), pp. 165–172.

Sandhaeger, F. *et al.* (2019) 'Monkey EEG links neuronal color and motion information across species and scales', *eLife*, 8, pp. 1–21. doi: 10.7554/eLife.45645.

Stokes, M. G., Wolff, M. J. and Spaak, E. (2015) 'Decoding Rich Spatial Information with High Temporal Resolution.', *Trends in cognitive sciences*. Elsevier Ltd, 19(11), pp. 636–8. doi: 10.1016/j.tics.2015.08.016.

Thielen, J. *et al.* (2019) 'Evidence for confounding eye movements under attempted fixation and active viewing in cognitive neuroscience', *Scientific Reports*, 9(1), pp. 1–8. doi: 10.1038/s41598-019-54018-z.

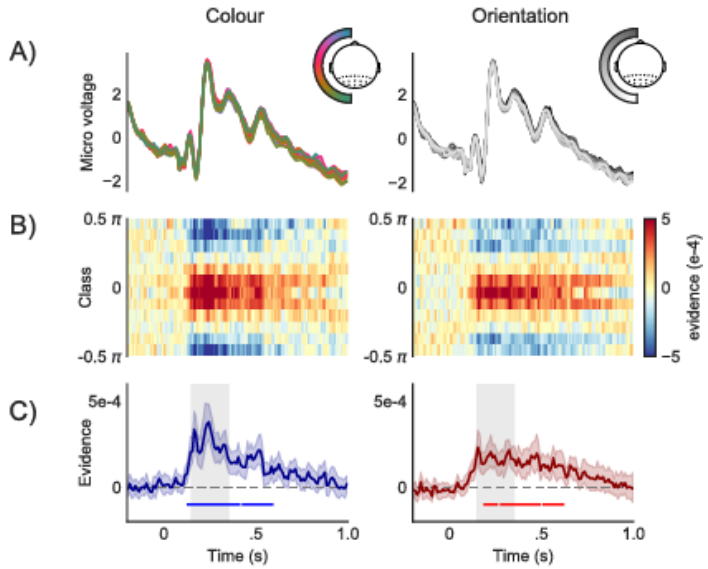
van Ede, F. *et al.* (2018) 'Decoding the influence of anticipatory states on visual perception in the presence of temporal distractors', *Nature Communications*. Springer US, 9(1), p. 1449. doi: 10.1038/s41467-018-03960-z.

van Ede, F. *et al.* (2019) 'Concurrent visual and motor selection during visual working memory guided action', *Nature Neuroscience*. Springer US, 22(3), pp. 477–483. doi: 10.1038/s41593-

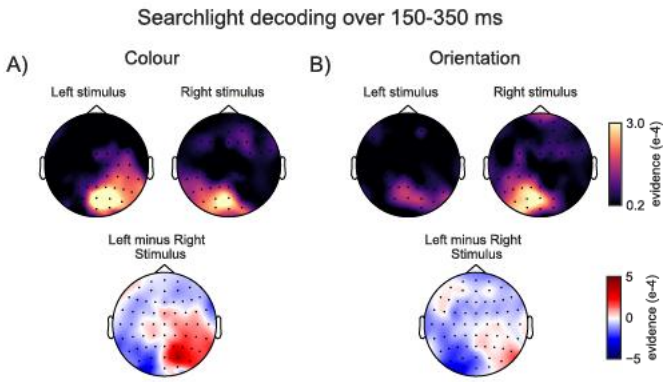
018-0335-6.
van Ede, F., Chekroud, S. R. and Nobre, A. C. (2019) 'Human gaze tracks attentional focusing in memorized visual space', *Nature Human Behaviour*. Springer US, 3(5), pp. 462–470. doi: 10.1038/s41562-019-0549-y.
Wachtler, T. and Wehrhahn, C. (1996) 'Perception of Luminance and Color', in Torre, V. and Conti, F. (eds) *Neurobiology: Ionic Channels, Neurons and the Brain*. Boston, MA: Springer US, pp. 383–389. doi: 10.1007/978-1-4615-5899-6_30.
Wolff, M. J. *et al.* (2017) 'Dynamic hidden states underlying working-memory-guided behavior.', *Nature neuroscience*, 20(6), pp. 864–871. doi: 10.1038/nn.4546.
Wolff, M. J. *et al.* (2019) 'Impulse responses reveal unimodal and bimodal access to visual and auditory working memory', *BioRxiv*, p. 623835. doi: 10.1101/623835.
Wolff, M. J. *et al.* (2020) 'Drifting codes within a stable coding scheme for working memory', *PLoS biology*, 18(3), p. e3000625. doi: 10.1371/journal.pbio.3000625.

Supplementary materials.

When applying LDA to the data from the 17 posterior electrodes for each time point, we do not reduce the dimensionality using PCA. Other preprocessing steps are identical. Single-time point LDA decoding shows that colour decoding was first significant after 125 ms, until 595 ms (**Supplementary Figure 1**). Orientation decoding was only significant after 185 ms, until 625 ms.



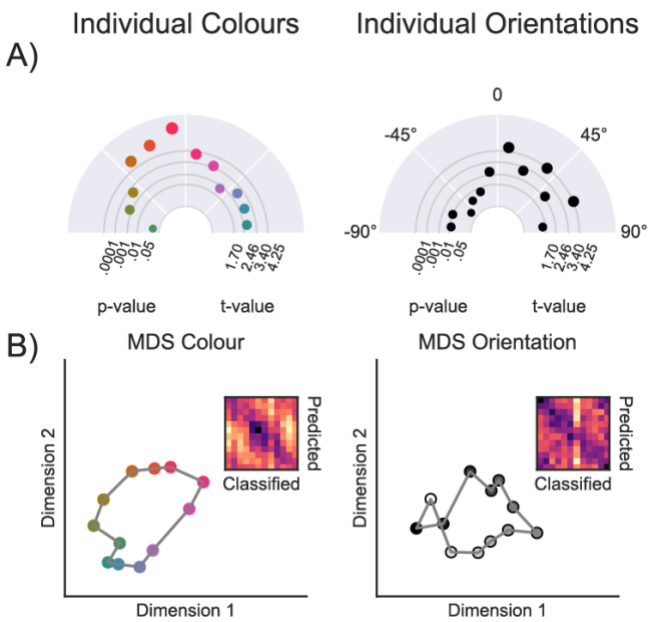
Supplementary Figure 1. Decoding performance for colours and orientations with single time-point decoding. A) Two dimensional tuning curves with the likelihood for target colour or orientation and all neighbouring values of theta. B) Mean cosine-convolved evidence with red lines for colour, blue for orientation. Cluster-permutation corrected significant time points are indicated with horizontal lines. Grey-shaded area between 150 to 350 ms is used for subsequent analyses. Error bars show 95% confidence intervals.



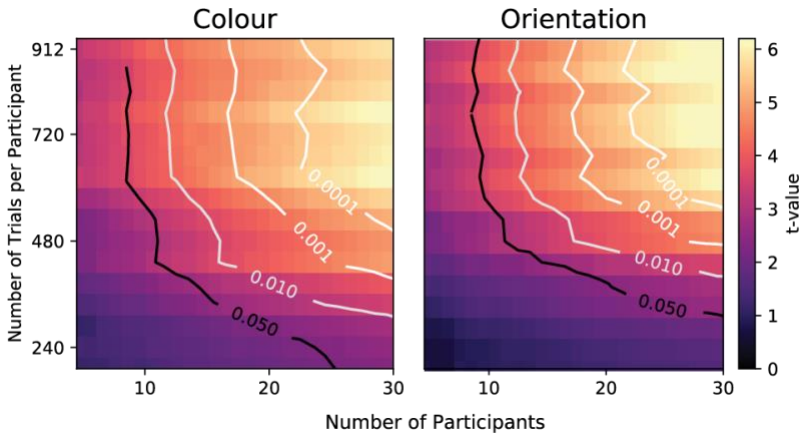
Supplementary figure S2. Colour and orientation decoding are primarily driven by posterior electrodes contralateral to the decoded stimulus. After averaging over data within the 150 to 350 ms time window, we ran a searchlight analysis. Top row topographies show which electrodes show highest mean cosine-convolved evidence for A) colours and B) orientations presented on the left and on the right. Posterior contralateral electrodes show highest evidence.

Bottom panels show the difference between left and right feature decoding topographies, highlighting the lateralisation of the evidence.

Instead of concatenating data over the window of 150 ms until 350 ms post stimulus onset, we averaged data into an average pattern and proceeded with the same analyses as depicted in (**Figure 4**). For colour bins, 11/12 colours could be decoded significantly against a $p < .05$ threshold (**Supplementary figure 3**; mean $t_{29} = 3.391$; min = 0.530; max = 6.040), and for orientation all 6/12 orientations could be decoded significantly against an uncorrected single-tail threshold (mean $t_{29} = 2.436$; min = 0.530; max = 4.561). After applying Bonferroni correction, 5/12 colour bins were significant compared to 4/12 orientations.



Supplementary figure 3. Decoding of individual features between 150 and 350 ms. A) This polar plot illustrates the decoding evidence of 12 individual colours and orientations. The distance from the centre depicts the t-value of testing the cosine-convolved tuning curve for a single feature, left and right combined, across participants, relative to zero. The size of the dots mirrors the magnitude of the evidence for that feature B) Two-dimensional visualisation of the similarity matrix in which we observed a circular configuration for colour features. Dissimilarity matrices show feature (colour or orientation) on the y-axis and distance between bins of the tuning curve on the x-axis.



Supplementary figure 4. Feature decoding as a function of trial and participant numbers. Average cosine-convolved evidence between 150 and 350 ms for the average of left and right feature decoding. Contours mark the .05, .01, .001, and .0001 p-value thresholds for n-1 degrees of freedom.