Seeing versus Knowing: The Temporal Dynamics of Real and Implied Colour Processing in the Human Brain

Abbreviated title: Real and Implied Colour Processing in the Brain

Lina Teichmann*[1,2], Tijl Grootswagers[1,2,3], Thomas Carlson[2,3], Anina N. Rich[1,2,4]
1| Perception in Action Research Centre & Department of Cognitive Science, Macquarie University, Sydney, Australia
2| ARC Centre of Excellence in Cognition & its Disorders, Macquarie University, Sydney, Australia
3| School of Psychology, University of Sydney, Australia
4| Centre for Elite Performance, Training and Expertise, Macquarie University, Sydney, Australia

Corresponding author's email: lina.teichmann@hdr.mq.edu.au

CONFLICT OF INTERESTS: none

ACKNOWLEDGEMENTS:
This research was supported by the Australian Research Council (ARC) Centre of Excellence in Cognition and its Disorders, International Macquarie University Research Training Program Scholarships to LT & TG, an ARC Future Fellowship (FT120100816) and an ARC Discovery project (DP160101300) to TC and an ARC Discovery project (DP170101840) to ANR.

Abstract

Colour is a defining feature of many objects, playing a crucial role in our ability to rapidly recognise things in the world around us and make categorical distinctions. For example, colour is a useful cue when distinguishing lemons from limes or blackberries from raspberries. That means our representation of many objects includes key colour-related information. The question addressed here is whether the neural representation activated by *knowing* that something is red is the same as that activated when we *actually see* something red, particularly in regard to timing. We addressed this question using neural timeseries (magnetoencephalography, MEG) data to contrast real colour perception and implied object colour activation. We applied multivariate pattern analysis (MVPA) to analyse the brain activation *patterns* evoked by colour accessed via real colour perception and implied colour activation. Applying MVPA to MEG data allows us here to focus, in particular, on the temporal dynamics of these processes. Male and female human participants (N=18) viewed isoluminant red and green shapes and grey-scale, luminance-matched pictures of fruits and vegetables that are red (e.g., tomato) or green (e.g., kiwifruit) in nature. We show that the brain activation pattern evoked by real colour perception is similar to implied colour activation, but that this pattern is instantiated at a later time. These results demonstrate that a common colour representation can be triggered by activating object representations from memory and perceiving colours. We show here that the key difference between these processes lies in the time it takes to access the common colour representation.

Significance Statement

Colour is a defining feature of objects and is a crucial cue to rapidly recognising and categorising things in our environment. In the current study, we recorded neural data to contrast the *timecourse* of colour representations when evoked by colour perception versus object-colour activation. We found that the neural representation for coloured shapes is similar to that of grey-scale images of objects that are associated with a colour (e.g., red for a tomato). However, the common colour representation is activated later when accessed via object-colour knowledge in comparison to colour perception, consistent with the idea that implied colours are evoked when the object representation is accessed. These results highlight the critical role of perceptual features such as colour in conceptual object representations.

Introduction

Throughout our lives, we learn statistical regularities about objects in our environment. We acquire knowledge about their typical perceptual features, which motor actions are required to interact with them, and in which context they usually appear. For example, we know that a tomato is round and red, we can eat it and it appears in the wider context of food. Our neural representations of objects therefore need to encompass a conceptual combination of these learnt attributes spanning from perception to action and semantic knowledge (Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995). The activation of object representations is likely to involve a widespread, distributed activation of several brain regions (Patterson, Nestor, & Rogers, 2007) with some brain areas responding preferentially to object colour (e.g., Seymour, Williams, & Rich, 2015). Several neuroimaging studies have compared perceiving colour and accessing object-colour knowledge from memory, finding evidence that similar brain areas are involved in these two processes (e.g., Bannert & Bartels, 2013; Martin et al., 1995; Vandenbroucke, Fahrenfort, Meuwese, Scholte, & Lamme, 2014). Using magnetoencephalography (MEG), we look at the neural timecourse of 'real' (by which we mean 'induced by wavelengths of light') colour perception versus implied object-colour activation from memory.

Associations between objects and typical or *implied colours* are acquired through experience (Bartleson, 1960; Hering, 1920) and are activated effortlessly and involuntarily (Bramão, Inácio, Faísca, Reis, & Petersson, 2010; Chiou, Sowman, Etchell, & Rich, 2014). The activation of object-colour knowledge is part of the dynamic interaction between perceptual processes and activation of prior conceptual knowledge to evaluate sensory input (Collins & Olson, 2014; Engel, Fries, & Singer, 2001; Goldstone, de Leeuw, & Landy, 2015). One of the central questions is how object-colour knowledge interacts or overlaps with colour representations generated by external stimuli. Is knowing that an object is typically a certain

colour the same as seeing that colour? There is behavioural evidence that object-colour knowledge can influence colour perception. (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006) found that participants overcompensated for implied colours when they were asked to change the colour of colour-diagnostic objects to be achromatic. For example, a banana would be adjusted towards the blue side of grey, showing the influence of the implied colour yellow. Similarly, (Witzel, 2016) showed that participants selected an image of an object as achromatic more often when its colour was modified to be the opposite of its implied colour (e.g., a bluish-grey banana). These results suggest that colour perception can be influenced by previously learnt object-colour associations (see Firestone & Scholl (2016) for debates about the extent to which activation of colour from memory is identical to colour perception). Brain-imaging data, recorded with functional magnetic resonance imaging (fMRI), suggest that brain activation corresponding to implied object colour activation shares characteristics with real colour perception: Retrieving the knowledge that a banana is yellow activates brain areas in or around the V4 complex, which is involved in colour perception (Bannert & Bartels, 2013; Barsalou, Simmons, Barbey, & Wilson, 2003; Chao & Martin, 1999; Rich et al., 2006; Simmons et al., 2007; Vandenbroucke et al., 2014). This is evidence that activation of implied colour rests on a similar neural architecture as real colour perception.

These results suggest that similar brain areas are active when perceiving colour and accessing implied colour, which may drive the behavioural interactions between the two (e.g., Hansen et al., 2006). Real colour activations occur very early in visual processing, whereas implied colour presumably is only activated once the object is processed at a higher level. Hence, there could be a temporal delay for activity driven by implied colour in comparison to activity driven by perceived colour. As the signal measured by fMRI is slow, it is not a suitable method to distinguish fine temporal differences between real and implied object colour processing. In the current study, we use multivariate pattern analysis (MVPA) on

MEG data (Grootswagers, Wardle, & Carlson, 2017) to compare the brain activation patterns evoked by colour perception and implied object colour activation. MEG has fine temporal resolution, and with MVPA we can detect patterns across the sensors at each time point that are reliable enough to train an algorithm to classify different categories of stimulus. Here, we use these methods to test whether a classifier trained on 'real colour' can successfully decode 'implied colour'. Such cross-generalisation can only occur if there is sufficient similiarity in the neural signals. This approach enables us to contrast the temporal dynamics of real and implied colour processing, shedding light on the interaction between perceptual processing and activation of object representations.

Materials and Method

Participants. 20 healthy volunteers (12 female, mean age = 27.6 years, SD = 6.6 years) completed the study. All participants reported normal or corrected-to-normal vision including normal colour vision. Participants gave informed consent before the experiment and were reimbursed with $20/hour. During the MEG recording, participants were asked to complete a target-detection task to ensure they were attentive. Two participants performed more than three standard deviations below the group mean on this task, suggesting they did not pay attention to the stimuli, and were therefore excluded from analysis, leaving 18 participants in total. The study was approved by the Macquarie University Human Research Ethics Committee.

Procedure. While lying in the magnetically shielded room (MSR) for MEG recordings, participants first completed a colour flicker task (Kaiser, 1991) to equate the coloured stimuli in perceptual luminance. Then they completed the main target-detection task. We used only two colours to increase the power of our analysis. We chose red and green as the two colours as they can be matched for luminance, and we included varying exemplars of these two hue categories to ensure any potential slight differences in luminance could not be used by a

classifier to distinguish the colour categories. While previous studies have shown the greatest behavioural effects for colours along the daylight axis (yellow-blue, (Hansen et al., 2006)), these are not feasible colours for the current design. If one seeks to make blue and yellow stimuli equiluminant, they are no longer perceived as blue and yellow. If there are luminance differences between the colour categories, the classifier can use this strong signal to discriminate the categories instead of relying on colour. In the colour flicker task, participants were presented with red and green circles (5x5 degrees visual angle) in the centre of the screen. The colours alternated at a rate of 30Hz. Participants completed 2 runs of 5 trials each. In each trial, one red-green combination was used. The red colour was kept consistent throughout each trial while participants were asked to use a button box to adjust the luminance level of green and report when they perceived the least amount of flickering. The HSV (hue, saturation, value) values for each green shade were then recorded. This procedure was repeated in the second run. The average HSV values between the two runs was then computed, yielding five shades of red and green equated for perceptual luminance. Using different shades of red and green which were each equated for perceptual luminance minimises the degree that any luminance difference between the categories could influence the results (see supplementary material for a table summarising individual HSV values used).

In the main target-detection task (Figure 1A), participants completed eight blocks of 440 trials each. There were two different types of blocks: *implied colour* and *real colour*. Block types alternated for each participant and the overall order was counterbalanced across participants. In the *implied colour* blocks, participants viewed luminance equated (SHINE toolbox, (Willenbockel et al., 2010)) grey-scale images of colour diagnostic objects (see Figure 1A). Equating the overall luminance ensures that differences in MEG response patterns are not caused by luminance differences between the red and green category. To increase variability in the stimulus set, half the stimuli depicted a single item on the screen (e.g., one strawberry)

and the other half were multiple, partially overlapping items (e.g., three strawberries). Having several different stimuli in each category helps to minimise the influence of low-level features such as edges and shapes on the results. In the *real colour* blocks, participants viewed five different abstract shapes. Each shape was filled in one of the red and green shades which were equated for perceptual luminance with the colour flicker task. Each shape occurred equally often in red and green. To match the stimuli presented in the *implied colour* block, half of the shapes were single shapes (e.g., one square) on the screen while the other half consisted of partially overlapping shapes (e.g., three squares). All stimuli (objects and shapes) contained the same number of pixels.

In both block types, presentation location varied randomly within 100 pixels around the central fixation cross. Changing the spatial location of the stimulus images adds variability to the retinal image, again reducing the influence low-level visual features have on the results. Participants were asked to press a button when they saw an image of the target shape (cross) or target object (capsicum). Every block had 40 target trials. All target trials were discarded before the MEG data analysis. On average, participants detected 98.51% (SD = 0.013%) of target stimuli.

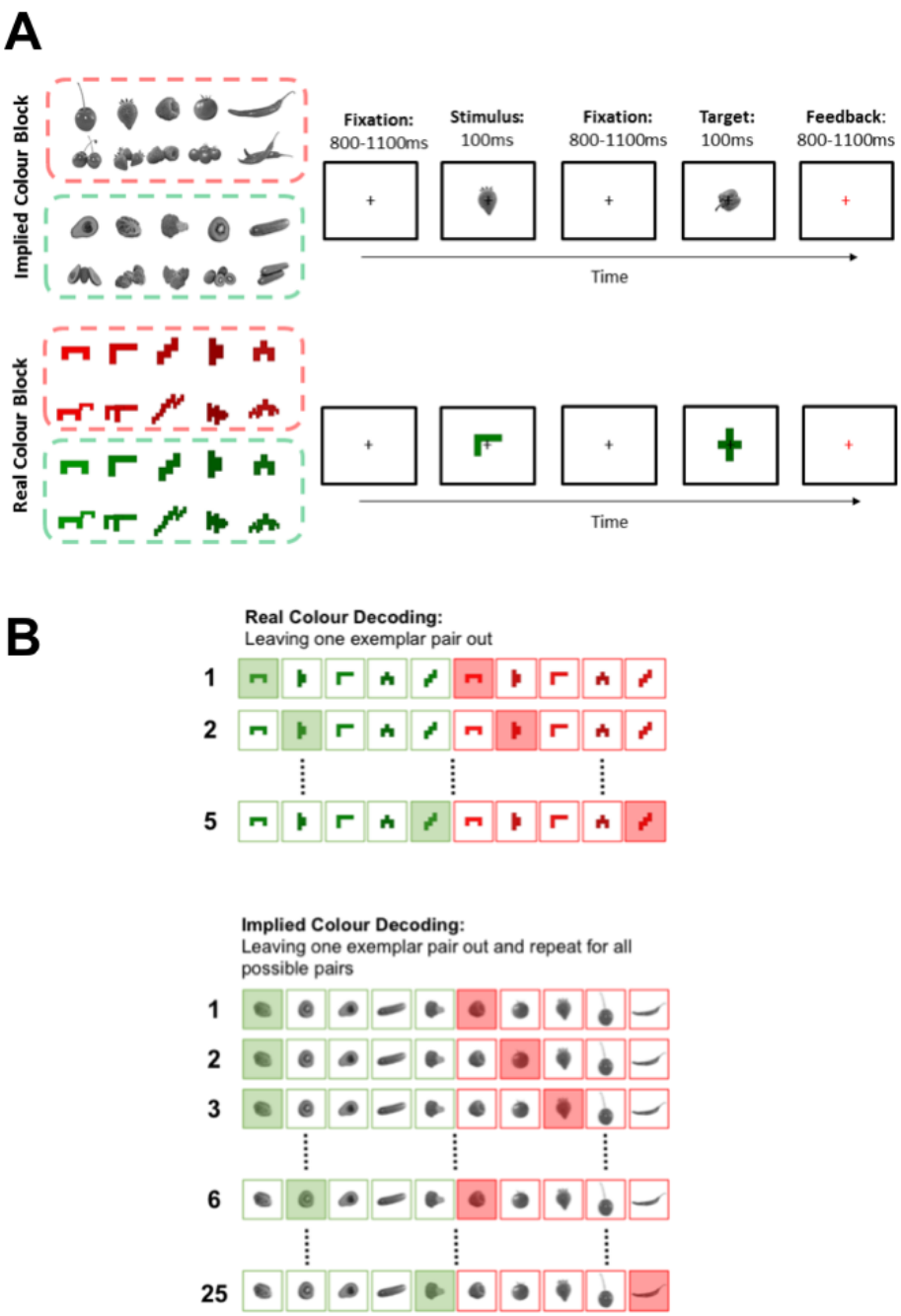

Figure 1: (A) Target-detection task and stimuli for both implied colour (top panel) and real colour (bottom panel) blocks. Participants were asked to press a button as soon as they saw the target (capsicum or cross). If they pressed the button after a target (correct detection) the fixation cross would turn green briefly, if they missed the target it would turn red. (B) Cross-validation approach used for real (top) and implied (bottom) colour decoding analyses. Every row shows which trials were used for the training set (clear) and which trials were used for the testing set (shaded). Trials with the same exemplar are never in the training and the testing set. In the real colour decoding analysis, we split the data in 5 different ways, always leaving one pair of the same shape with matched luminance out. In the implied colour decoding analysis, we split the data in 25 different ways, leaving all possible exemplar pairs out once. The classification accuracy is an average of the classifier performance for each of these divisions.

Apparatus and Pre-processing. The stimuli had a visual angle of approximately 5 x 5 degrees and were projected onto a translucent screen mounted in the magnetically shielded room. Stimuli were presented using MATLAB with Psychtoolbox extension (Brainard & Pelli, 1997; Kleiner et al., 2007; Pelli, 1997). The neuromagnetic recordings were obtained with a whole-head axial gradiometer MEG (KIT, Kanazawa, Japan), containing 160 axial gradiometers. The frequency of recording was 1000Hz. FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) was used to pre-process the data. We used a low-pass filter of 200Hz and a high-pass filter of 0.03Hz online. Stimulus onsets were determined with a photodiode that detected light change when a stimulus came on the screen. Trials were epoched from -100 to 800ms relative to stimulus onset and downsampled to 200Hz (5ms resolution). All target trials were removed. Then principal component analysis (PCA) was conducted across the remaining trials to reduce dimensionality and separate noise from the data. We calculated the cumulative sum of variance explained by each component and retained components until 99% of the variance in the data could be explained. Consequently, there was a different number of components at every timepoint. We performed no further preprocessing steps (e.g., channel selection, artefact correction).

Decoding Analysis. We conducted four separate decoding analyses using linear discriminant classifiers (LDA). CosmoMVPA (Oosterhof, Connolly, & Haxby, 2016) was used to conduct these analyses. First, to test whether we can decode perception of red versus green, we analysed the data from the real colour (shape) blocks. We tested whether we could decode the colour of our abstract shapes for each person. The classifier was trained on distinguishing the activity patterns evoked by red versus green shapes at each timepoint using 80% of the data. We then used the classifier to predict the colour of each stimulus at every timepoint in the remaining 20% of the data. To divide the data into training and testing set, we left out one exemplar pair with matched luminance (e.g., red and green L-shape, matched for perceptual

luminance). This process was repeated over all folds so that each exemplar pair was in the training and the testing set once (5-fold cross-validation). Hence, the colours in each fold were balanced (Figure 1B).

Second, to assess whether we can decode implied colour from grey-scale objects, we trained a classifier to distinguish trials of grey-scale objects that are associated with red versus green. We used an independent exemplar cross-validation approach (Carlson, Tovar, Alink, & Kriegeskorte, 2013) for the implied colour decoding analysis to rule out that the classification is driven by visual features such as shape. We selected trials based on label for both colour categories (e.g., all strawberry and kiwifruit trials). We then trained our classifier to distinguish between activity patterns evoked by all stimuli *except* the selected stimuli. The selected trials were then used as testing set. Thus, the classifier did *not* train on any trials of the test objects, ensuring that the classifier cannot rely on specific shapes. We repeated this process 25 times to have every possible combination of green and red objects used once as the testing set, and report the average classification performance over all these combinations (Figure 1B).

Last, we conducted a cross-decoding analysis across the two different block types, training the classifier on all real colour trials and testing on all implied colour trials. This cross-decoding analysis is highly conservative as everything about the stimuli differs between real colour and object colour trials, the only potential link is the implied colour of the objects to the real colour of the abstract shapes. It is, however, quite likely that accessing the colour information about objects takes longer than perceiving colour directly. This means that even if a similar pattern is elicited by the two colour types, we may not see it in a direct cross-decoding analysis. We therefore also conducted a time-generalisation analysis (Carlson, Hogendoorn, Kanai, Mesik, & Turret, 2011; King & Dehaene, 2014), training the classifier at each timepoint on the real colour trials and then testing on each timepoint in implied colour trials, yielding an accuracy score for all pairs of training and testing timepoints. This technique allows us to test

for similar activation patterns that do not occur at the same time. As a different number of principal components was retained at every timepoint, we ran the time-generalisation analysis on the raw sensor activations instead of the principal components.

To assess whether decoding accuracy was significantly above chance, in all analyses, we used random effects Monte-Carlo cluster statistic using Threshold Free Cluster Enhancement (TFCE, (Smith & Nichols, 2009)) as cluster-forming statistic, as implemented in CosmoMVPA (Oosterhof et al., 2016). The TFCE statistic represents the support from neighbouring time points, allowing optimal detection of sharp peaks, as well as sustained weaker effects. We conducted a permutation test to generate null distributions of the TFCE statistic. For each subject, the colour labels at each timepoint were randomly shuffled 100 times and the decoding analysis was repeated using the shuffled labels. To keep the temporal structure, the same label permutations were applied to all time points in the data. Next, a group-level null distribution was constructed by randomly drawing from the subject null-distributions 10,000 times (1000 times for the time-generalisation analysis). Then, to correct for multiple comparisons, the most extreme value of the null distribution over time points was taken in order to construct an overall null distribution across the time-series. The 95th percentile of this corrected null distribution was used to threshold the observed decoding results for significance.

Results

For our real colour decoding analysis, we trained the classifier to distinguish red from green shapes and then tested whether it could distinguish red from green shapes in an independent set of data. The classifier was able to predict the colour above chance in a cluster stretching from 70 to 250 ms, reflecting a signal modulated by colour (Figure 2A).

To examine whether we can decode implied object colour, the classifier was trained on a subset of the object trials and then tested on an independent set. The testing set included only

exemplars (e.g., all strawberry and kiwifruit trials) that the classifier did not train on. The implied colour decoding was not significant at any point in the timeseries (Figure 2B).

To test whether there is a representational overlap of real and object colour processing we ran a cross-decoding analysis. For this analysis, we trained the classifier to distinguish between the red and green shapes and tested its performance on the grey-scale objects to see whether cross-generalisation between real and implied object colour is possible. In this analysis, the classifier is trained on identical shapes that only vary in terms of colour. Hence, this is the most conservative way of testing whether there is a representational overlap between real and implied colours. The cross-colour decoding was not significant at any point in the timeseries (Figure 2C)

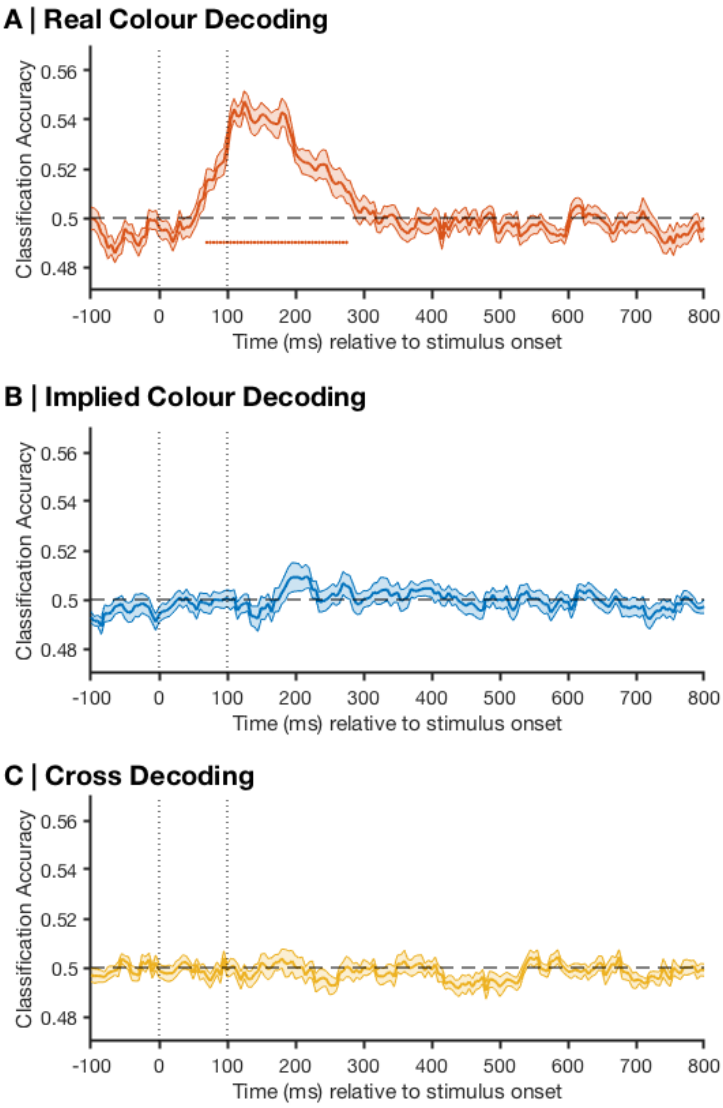

Figure 2: Classification accuracies for (A) real colour, (B) implied colour, and (C) cross-colour decoding over time. Vertical, dotted lines show stimulus on- and offset. Dashed line indicates chance level (50%). Shading indicates error bars around the mean (standard deviation of decoding accuracies across participants divided by the square root of the number of participants). Coloured dots depict significant timepoints corrected for multiple comparisons.

Accessing implied colour presumably requires first accessing the general concept of the object. Therefore, it is possible that real and implied colours are represented in a similar way but that colour information is accessed later when activated via objects in comparison to when colour is perceived. To examine whether this is the case we ran a cross-decoding time-generalisation analysis. We trained a classifier to distinguish between red and green shapes at every timepoint and then tested whether it could cross-generalise to the grey-scale objects at any timepoint. The results of the cross-generalisation analysis are summarised in Figure 3.

The time-generalisation analysis revealed similar activation patterns between real and implied colours when the classifier is trained on real colour at an earlier timepoint and tested on implied colour at a later one (Figure 3A and 3B). These generalisation accuracies were statistically above chance (Figure 3C). Inspecting the training timepoint with maximum decoding (135 ms) indicates that there is above-chance decoding at later testing timepoints with peak decoding at 195 ms after stimulus onset (Figure 3B). The results show that we can cross-decode between real and implied colours when we *train* the classifier on *real* colours at timepoints between 125 to 155 ms and *test* it on *implied* colours at a cluster from 185 to 230 ms (Figure 3E). We interpret this main cluster of significant decoding, as it is more than three times larger than the other significant timepoints combined, highlighting that these are the timepoint combinations with stable decoding. Combining the off-diagonal shift of *all* significant timepoints shows a median delay of 57.5 ms for implied colour testing times compared to real colour training times (Figure 3D, 3F). Importantly, these results cannot be driven by anything else than colour as the classifier is trained on real colour trials in which the only different stimulus characteristic was colour and tested on implied colour trials which were achromatic. The results highlight that there are similarities between real colour and implied object colour patterns but this pattern is instantiated later for implied object colours than for real colours. Note that above-chance cross-decoding does not mean that the processes involved

in real and implied colour processing are identical. However, the results show that there are sufficient similarities for the classifier to cross-generalise from brain activation patterns evoked by perceiving red and green to brain activation patterns evoked by viewing grey-scale images of objects that are associated with red and green.

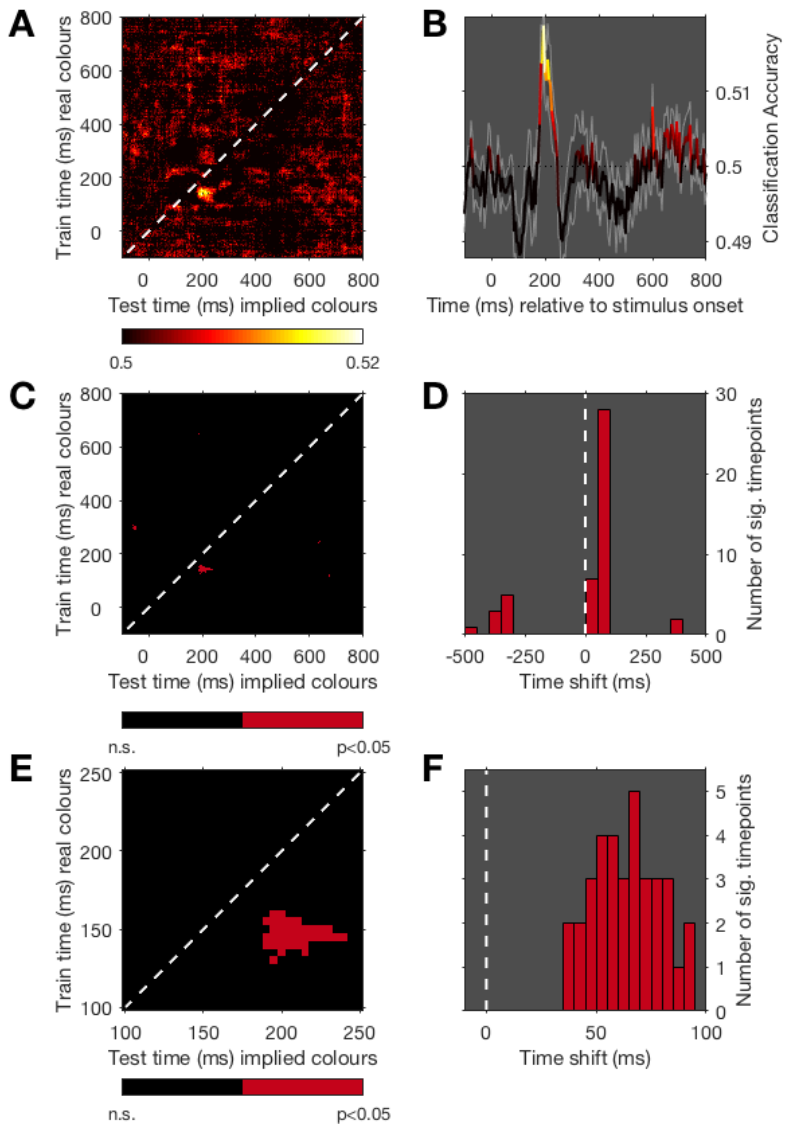

Figure 3: Results of the time-generalisation analysis. In this analysis, the classifier was trained on the real colour trials and tested on the implied colour trials. Panel A shows the classification accuracy at every possible train-test-time combination. Lighter colours indicate higher classification accuracy. Panel B shows the timecourse of the classification accuracy when the classifier relies on the training data at 135ms (peak decoding at 195ms). Shading indicates error bar around the mean (standard deviation of decoding accuracies across participants divided by the square root of the number of participants). Panel C shows all training-testing-timepoint combinations where classification is significantly above chance (based on permutation, corrected for multiple comparisons). Panel D shows the time shift of all significant timepoints from the diagonal. Most significant timepoint combinations (35 out of 47) fall into a cluster reflecting a delay of ~ 35 to 95ms. Panel E and F are zoomed in versions of panels C and D for this cluster, highlighting that there is an offset of around 70ms for the significant timepoints within the cluster.

These results predict that it takes more time to access implied colour than real colour as one presumably has to first access the concept of the object. We decided post-hoc to test this prediction and collected colour categorisation accuracies and reaction times on our stimuli from a new sample of 100 participants on Amazon's Mechanical Turk. Participants were presented with the red and green shapes and the grey-scale objects, each presented individually for 100ms, and randomly intermingled. Participants were asked to indicate as quickly and accurately as possible whether the stimulus was (typically) red or green by responding with either "m" or "z" using a keyboard. This allowed us to first confirm that the objects we had selected were indeed typically associated with red or green, and second, to test whether there was a time difference between real and implied colour categorisation. Response-key mappings were randomly determined for each participant. Four participants were excluded from the analysis as their accuracy scores were more than 2 standard deviations below the group mean. For the remaining 96 participants, choosing which colour is displayed was faster for the coloured shapes (M=506.74ms, 95% CI [491.71, 521.77]) than indicating the implied colour of the grey-scale images (M=607.15ms, 95% CI [589.98, 624.31]). Real colour responses were also more accurate (M=91.57%, 95% CI [90.89, 92.25]) than implied colour responses (M=80.32%, 95% CI [79.37, 81.26]). Using Mturk introduces variance to the experimental setup, including monitor settings for colour, computer and internet speeds, all of which will increase the noise in the data; we do not, therefore, interpret the specific difference in timing. Despite the variability, there is a clear difference between the time taken for categorising colour in the two conditions. These results are consistent with real colour perception being faster and easier than recalling implied colours, in line with the prediction from our decoding results.

Discussion

In this study, we compared the temporal activation profiles of colour perception and implied colour to examine the interaction between perceptual processing and access to object representations. We applied MVPA to time-series MEG data and show that real and implied colour processing share a sufficient degree of similarity to allow for cross-generalisation with a temporal shift. The activity pattern distinguishing colours was instantiated on average 57ms later for implied colours than for real colour, highlighting that there are similarities between colour representations accessed via 'real' colour and via implied colour, but that there is a temporal asynchrony between these processes.

There are several possible explanations for a temporal difference between accessing colour representations via real colour and implied colour. One possibility is that the time difference reflects greater individual variability in the temporal activation profile of implied colours in comparison to real colours. Implied colours may be accessed at slightly different timepoints for different people and thus the cross-decoding accuracy that is above chance for each participant only overlaps at a later timepoint. There are also alternative, more interesting, potential explanations. First, it could be due to actual differences in neural processes. Colour representations accessed via colour perception are immediately available whereas implied colour activation presumably only happens once the object is processed to some higher level. Thus, the delay could reflect differences between bottom-up and top-down access to colour representations. This is a plausible interpretation of the temporal delay and corresponds with earlier findings of early visual areas being involved in implied colour activation (Bannert & Bartels, 2013; Vandenbroucke et al., 2014). Second, it is possible that the binding of colour and shape information happens later in the visual processing hierarchy compared to initial processing of separate features, and that the comparison of typical and perceived colour can therefore only happen later once shape-colour binding is complete. This view is consistent with

results of a recent fMRI study, which showed that object-colour and object-shape activated from memory can be distinguished in areas associated with colour (V4) and shape (lateral occipital cortex, LOC) perception, respectively, but that the conjunction of colour *and* shape can be decoded only later along the visual hierarchy (anterior temporal lobe, ATL; (Coutanche & Thompson-Schill, 2014)). Similarly, (Seymour et al., 2015) showed that colour *per se* can be decoded in early visual areas but object surface colour (bound to form) can only be decoded in areas further along the ventral visual stream. These findings also correspond to patient work (Patterson et al., 2007) and previous transcranial magnetic stimulation studies (Chiou et al., 2014) which point towards the ATL as the hub for object-knowledge (for a recent review see Lambon Ralph, Jefferies, Patterson, & Rogers, 2017). Finally, it could also be that the delay reflects the greater complexity of the grey-scale objects relative to the abstract shapes, hence binding the features may take slightly longer. From the data we cannot disentangle these interpretations. Our results clearly highlight, however, that there is a similar structure to the brain response to externally perceived and internally activated colour representations, and that time seems to be the key difference.

We interpret our cross-decoding results as evidence that representation of implied colour involves some of the same mechanisms as those involved in colour perception. This is in line with previous studies showing that the same brain regions are active when retrieving object representations from memory and perceiving those object features (for reviews see Martin, 2007; Patterson et al., 2007). For example, (Vandenbroucke et al., 2014) and (Bannert & Bartels, 2013) showed that early visual cortex is involved when real and implied colour are processed. Using fMRI, (Vandenbroucke et al., 2014) trained a classifier on data recorded while participants viewed red and green shapes, and then tested the classifier on data recorded while participants viewed line-drawings of colour-diagnostic objects filled with ambiguous colours between red and green. Behavioural results suggested participants were slightly more

likely to respond ‘red’ to the ambiguous colour presented on a line drawing of a typically red object than a line drawing of a typically green object. In their fMRI data, the classifier categorised the colour consistent with what the participant perceived. That means the classifier categorised colours to be red when shown on objects that are typically red and green for objects that are typically green at above chance levels. They interpret these data as evidence for an influence of implied object colours on the creation of a subjective experience of colour. Consistent with this study, (Bannert & Bartels, 2013) trained a classifier to distinguish fMRI data from trials where four different colours were presented. They showed that the classifier can cross-generalise to grey-scale colour-diagnostic objects. Both fMRI studies highlight that there are similar activation patterns across voxels in the visual cortex for real and implied colour processing. Our results provide further evidence that object-colour knowledge and colour perception instantiate similar patterns, this time in the temporal domain.

What is driving the successful decoding performance? Here, we decoded real and implied object colour. For the real colour decoding, we used shapes that were identical across colour categories and used five different levels of stimulus luminance for each category that were perceptually matched. Therefore, the only distinguishing feature between the shapes was colour. That means that for the real-colour decoding analysis *and* the cross-generalisation (i.e., training on shapes and testing on objects), we can rule out visual differences other than colour as a driving factor. For the implied colour blocks, visual differences are a bigger concern. Previous studies have used line-drawings instead of photos of objects to reduce local low-level differences between stimuli (e.g., Vandenbroucke et al., 2014). Line-drawings can reduce some of these differences (e.g., local luminance differences due to texture) but cannot completely rule out any contribution of low-level effects (e.g., shapes). In addition, there is a considerable trade-off between line-drawings in terms of similarity of the objects to real world objects which can slow down recognition and implied colour effects (Olkkonen, Hansen, & Gegenfurtner,

2008; Vurro, Ling, & Hulbert, 2013). Visual differences, however, are not a concern for the key cross-decoding analysis. Here, we used identical shapes in the training set, making low-level features a highly unlikely source of contribution to classifier performance. These controls allow us to infer that the classifier is using (implied) colour to distinguish the categories.

There may be several reasons why decoding implied colour was unsuccessful when training the classifier on distinguishing within grey-scale objects and yet successful for the cross decoding. First, for the cross-decoding analysis, the classifier must learn to distinguish between colours alone – there is no other cue that can be used to complete the task as all other differences are controlled (shape, luminance, texture, edges, etc). In comparison, when the classifier trains on the grey-scale objects in the implied colour trials, there are stimulus differences in addition to implied colour that are likely to be much stronger signals; this results in overfitting to visual differences such as shape or texture which are then not useful in the test on implied colour of different objects. We limited the impact of these differences on decoding accuracy by leaving exemplar pairs out when training the classifier, but the classifier may nevertheless overfit as these differences are much stronger than any implied colour. Thus, when training on the shape trials (only colour differences), we can be sure the classifier is learning to differentiate *colours* from brain activation patterns rather than any other signals. In contrast, we cannot be certain that this is happening when we train on greyscale objects that are associated with red and green. Second, the cross-decoding analysis does not require any cross-validation as the trials are independent. That means the cross-decoding analysis has more power than the implied-implied decoding analysis. Note, there is a small peak in the implied-implied decoding at around 200ms after stimulus onset that did not survive correction; it is possible this reflects an effect we cannot detect due to lack of power and our conservative correction. In the cross-decoding analysis we train and test on 20% more data and therefore have more power to detect differences.

Our results speak to the important aspect of temporal differences between colour evoked by external stimulation and internal activation. Activating conceptual knowledge of objects from memory is thought to involve a widespread network of brain areas involved in perception and action of different object features (Martin, 2007; Patterson et al., 2007). To access the implied colour of an object requires that the conceptual representation of that object is activated first. Using time-generalisation methods (King & Dehaene, 2014), we show here that in comparison to real colour perception, which can be decoded rapidly, accessing object-colour knowledge takes on average ~57ms longer. This is consistent with our behavioural data showing that real colour judgments are faster than implied colour judgments. These results increase our existing knowledge of how real and implied colour are processed by showing that colour representations via external stimulation are also instantiated during internal activation, but with a delay. Applying MVPA to our MEG data allows us to capture this temporal difference, highlighting the value of this method for dissociating activation of object features from memory from perception of object features in the real world.

Our results highlight that the activation of implied colours can occur independent of a task that focuses on colour. Participants completed a target-detection task in which attending to colour was not a useful strategy. The targets were ambiguous in colour (e.g., a capsicum can be either red or green), and this avoided biasing participants towards deliberately thinking about the implied colour of the objects. Using a task that is irrelevant to the classifier performance allowed us to explore the involuntary activation of implied colours rather than the signals associated with perhaps actively imagining colours or retrieving colour names. Not attending to the feature that is relevant for the classifier probably reduced our decoding accuracy in general (e.g., Brouwer & Heeger, 2013; Jackson, Rich, Williams, & Woolgar, 2016), but clearly supports previous studies showing that there is an *involuntary* activation of

object-colour independent of task demands (Bannert & Bartels, 2013; Vandenbroucke et al., 2014).

Overall, the decoding accuracies across our analyses are quite low. As outlined above, this is probably partially due to colour being irrelevant for the task. In addition, it is important to note that we did not use extensive pre-processing, meaning we ran our analyses on effectively raw data. We use our multivariate decoding analyses for *interpretation* (Hebart & Baker, 2017) – if decoding is above chance, this means there is a signal that allows a categorical distinction between the conditions. Minimal pre-processing (e.g., no trial averaging, filtering, channel-selection, trial-selection) ensures that there is no potential influence of plurality of methods or specific pre-processing choices; it also means that the data overall are noisier which can result in relatively low decoding accuracies. However, it is crucial to note that low decoding accuracies does not necessarily mean that the effects are weak, as decoding accuracies are not effect sizes (cf. Hebart & Baker, 2017). We recently replicated this main finding using a different task, stimuli and timings (Teichmann, Quek, Robinson, Grootswagers, Carlson, Rich, in preparation). Despite these differences, there was significant cross-generalisation from 'real colour' to 'implied colour' across a similar cluster of time-points. The magnitude of the decoding accuracy is less important than the fact it is possible. Here, we show with rigorous methodological controls and strict correction for multiple comparisons that there is significant cross-generalisation from real colour to implied colour.

Previous fMRI studies showed that early visual areas are involved in real colour perception and implied colour activation (Bannert & Bartels, 2013; Rich et al., 2006; Vandenbroucke et al., 2014), but other studies implicate anterior temporal regions in object colour knowledge. For example, a transcranial magnetic stimulation study showed that the behavioural effects of implied colour knowledge on object recognition are disrupted when stimulating the anterior temporal lobe (Chiou et al., 2014), complementing patient studies

suggesting this area holds conceptual object information (e.g., Ralph, Matthew, & Patterson, 2008). This highlights that activating object attributes, including implied colour, goes beyond low-level visual areas. Our study adds time as a novel aspect to this discussion by comparing the temporal profiles of colour representations accessed via real colour perception and implied colour activation.

In conclusion, our data show that there is a common representation of real and implied colour but that this representation is accessed later when triggered by activating implied colour than by perceiving real colour. This is in line with previous studies suggesting that the same brain areas are involved in object-feature activation from memory *and* object-feature perception. Our results highlight that applying MVPA to time-series MEG data is a valuable approach to exploring the interaction between object-feature inputs and predictions or representations based on prior knowledge. This opens multiple avenues for future studies examining the dynamic interactions between perceptual processes and activation of prior conceptual knowledge.

References

Bannert, M. M., & Bartels, A. (2013). Decoding the yellow of a gray banana. *Current Biology, 23*(22), 2268-2272.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in cognitive sciences, 7*(2), 84-91.

Bartleson, C. J. (1960). Memory colors of familiar objects. *JOSA, 50*(1), 73-77.

Brainard, D. H., & Pelli, D. G. (1997). The psychophysics toolbox. *Spatial vision, 10*, 433-436.

Bramão, I., Inácio, F., Faísca, L., Reis, A., & Petersson, K. M. (2010). The influence of color information on the recognition of color diagnostic and noncolor diagnostic objects. *The Journal of general psychology, 138*(1), 49-65.

Brouwer, G. J., & Heeger, D. J. (2013). Categorical clustering of the neural representation of color. *Journal of Neuroscience, 33*(39), 15454-15465.

Carlson, T., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution decoding of object position and category. *Journal of vision, 11*(10), 9-9.

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: the first 1000 ms. *Journal of vision, 13*(10), 1-1.

Chao, L. L., & Martin, A. (1999). Cortical regions associated with perceiving, naming, and knowing about colors. *Journal of cognitive neuroscience, 11*(1), 25-35.

Chiou, R., Sowman, P. F., Etchell, A. C., & Rich, A. N. (2014). A conceptual lemon: theta burst stimulation to the left anterior temporal lobe untangles object representation and its canonical color. *Journal of cognitive neuroscience, 26*(5), 1066-1074.

Collins, J. A., & Olson, I. R. (2014). Knowledge is power: How conceptual knowledge transforms visual cognition. *Psychonomic bulletin & review, 21*(4), 843-860.

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Creating concepts from converging features in human cortex. *Cerebral cortex, 25*(9), 2584-2593.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience, 2*(10), 704-716.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and brain sciences, 39*.

Goldstone, R. L., de Leeuw, J. R., & Landy, D. H. (2015). Fitting perception in and to cognition. *Cognition, 135*, 24-29.

Grootswagers, T., Wardle, S. G., & Carlson, T. 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*.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature neuroscience, 9*(11), 1367-1368.

Hebart, M. N., & Baker, C. I. (2017). Deconstructing multivariate decoding for the study of brain function. *Neuroimage*.

Hering, E. (1920). *Grundzüge der Lehre vom Lichtsinn*: Springer.

Jackson, J., Rich, A. N., Williams, M. A., & Woolgar, A. (2016). Feature-selective Attention in Fronto-parietal Cortex: Multivoxel Codes Adjust to Prioritize Task-relevant Information. *Journal of cognitive neuroscience*.

Kaiser, P. (1991). Flicker as a function of wavelength and heterochromatic flicker photometry. *Limits of vision (Kulikowski JJ, Walsh V, Murray IJ, eds)*, 171-190.

King, J., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in cognitive sciences, 18*(4), 203-210.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*(14), 1.

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience, 18*(1), 42.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol., 58*, 25-45.

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science, 270*(5233), 102.

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of vision, 8*(5), 13-13.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience, 2011*, 1.

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Frontiers in neuroinformatics, 10*.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*(12), 976-987.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision, 10*(4), 437-442.

Ralph, L., Matthew, A., & Patterson, K. (2008). Generalization and differentiation in semantic memory. *Annals of the New York Academy of Sciences, 1124*(1), 61-76.

Rich, A. N., Williams, M. A., Puce, A., Syngeniotis, A., Howard, M. A., McGlone, F., & Mattingley, J. B. J. N. (2006). Neural correlates of imagined and synaesthetic colours. *44*(14), 2918-2925.

Seymour, K., Williams, M., & Rich, A. (2015). The representation of color across the human visual cortex: distinguishing chromatic signals contributing to object form versus surface color. *Cerebral cortex, 26*(5), 1997-2005.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia, 45*(12), 2802-2810.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage, 44*(1), 83-98.

Vandenbroucke, A., Fahrenfort, J., Meuwese, J., Scholte, H., & Lamme, V. (2014). Prior knowledge about objects determines neural color representation in human visual cortex. *Cerebral cortex, 26*(4), 1401-1408.

Vurro, M., Ling, Y., & Hulbert, A. (2013). Memory color of natural familiar objects: Effects of surface texture and 3-D.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior research methods, 42*(3), 671-684.

Witzel, C. (2016). An Easy Way to Show Memory Color Effects. *i-Perception, 7*(5), 2041669516663751. doi:10.1177/2041669516663751

Supplementary Material

Table 1: HSV values for the colours used in the real experiment. Five consistent levels of luminance were used for red while green was perceptually matched to the reds for each participant with a colour flicker task. Hue and saturation were kept consistent throughout.

Red: Five consistent levels of luminance

Red1			Red2			Red3			Red4			Red5		
H	S	V	H	S	V	H	S	V	H	S	V	H	S	V
0.00	0.90	0.70	0.00	0.90	0.77	0.00	0.90	0.84	0.00	0.90	0.91	0.00	0.90	0.98

Green: Perceptually matched per participant

	Green1			Green2			Green3			Green4			Green5		
Participant	H	S	V	H	S	V	H	S	V	H	S	V	H	S	V
1	0.333	0.9	0.405	0.333	0.9	0.435	0.333	0.9	0.48	0.333	0.9	0.51	0.333	0.9	0.505
2	0.333	0.9	0.405	0.333	0.9	0.455	0.333	0.9	0.48	0.333	0.9	0.47	0.333	0.9	0.52
3	0.333	0.9	0.405	0.333	0.9	0.425	0.333	0.9	0.45	0.333	0.9	0.495	0.333	0.9	0.49
4	0.333	0.9	0.4	0.333	0.9	0.475	0.333	0.9	0.39	0.333	0.9	0.455	0.333	0.9	0.475
5	0.333	0.9	0.44	0.333	0.9	0.475	0.333	0.9	0.485	0.333	0.9	0.52	0.333	0.9	0.56
6	0.333	0.9	0.525	0.333	0.9	0.525	0.333	0.9	0.53	0.333	0.9	0.535	0.333	0.9	0.56
7	0.333	0.9	0.39	0.333	0.9	0.415	0.333	0.9	0.42	0.333	0.9	0.475	0.333	0.9	0.495
8	0.333	0.9	0.385	0.333	0.9	0.415	0.333	0.9	0.515	0.333	0.9	0.49	0.333	0.9	0.485
9	0.333	0.9	0.415	0.333	0.9	0.43	0.333	0.9	0.445	0.333	0.9	0.475	0.333	0.9	0.48
10	0.333	0.9	0.395	0.333	0.9	0.415	0.333	0.9	0.435	0.333	0.9	0.47	0.333	0.9	0.45
11	0.333	0.9	0.45	0.333	0.9	0.43	0.333	0.9	0.495	0.333	0.9	0.485	0.333	0.9	0.47
12	0.333	0.9	0.42	0.333	0.9	0.455	0.333	0.9	0.48	0.333	0.9	0.51	0.333	0.9	0.55
13	0.333	0.9	0.42	0.333	0.9	0.425	0.333	0.9	0.415	0.333	0.9	0.455	0.333	0.9	0.455
14	0.333	0.9	0.5	0.333	0.9	0.47	0.333	0.9	0.48	0.333	0.9	0.505	0.333	0.9	0.565

15	0.333	0.9	0.46	0.333	0.9	0.445	0.333	0.9	0.485	0.333	0.9	0.525	0.333	0.9	0.52
16	0.333	0.9	0.455	0.333	0.9	0.52	0.333	0.9	0.495	0.333	0.9	0.54	0.333	0.9	0.545
17	0.333	0.9	0.41	0.333	0.9	0.445	0.333	0.9	0.43	0.333	0.9	0.47	0.333	0.9	0.475
18	0.333	0.9	0.43	0.333	0.9	0.455	0.333	0.9	0.575	0.333	0.9	0.615	0.333	0.9	0.57
MEAN	**0.33**	**0.90**	**0.43**	**0.33**	**0.90**	**0.45**	**0.33**	**0.90**	**0.47**	**0.33**	**0.90**	**0.50**	**0.33**	**0.90**	**0.51**