Published in final edited form as:

Cortex. 2019 February ; 111: 196–209. doi:10.1016/j.cortex.2018.10.013.

The time course of encoding and maintenance of task-relevant versus irrelevant object features in working memory

Andrea Bocincova¹ and Jeffrey S. Johnson¹

¹Department of Psychology and Center for Visual and Cognitive Neuroscience, North Dakota State University

Abstract

Access to WM can be restricted on the basis of goal-relevant properties such as spatial location. However, the extent of voluntary control over which features of an attended multi-feature object are encoded and maintained in WM is debated. Some evidence suggests that attending to an object leads to obligatory storage of all of its features, whereas other evidence suggests that access to WM can be restricted to only goal-relevant features. Another possibility is that all features are initially encoded, but irrelevant features are removed from WM over time. To address these various possibilities, we used pattern classification of EEG signals to track the temporal evolution of representations reflecting the encoding and storage of task-relevant and irrelevant features in WM. In different blocks, participants remembered the orientation, color or both orientation and color of a colored, oriented grating. The color and orientation of the grating was randomly drawn from two distinct feature bins on each trial. To examine trial-specific activity reflecting storage of the object's features, a support vector machine (SVM) classifier was trained to classify what bin the stimulus features came from. Importantly, for orientation, the classifier produced reliably above-chance classification across the delay when orientation was task-relevant but not when it was task-irrelevant. Interestingly, orientation could be accurately classified on trials for which both orientation and color were remembered. Moreover, a separate measure corresponding to the probability of a feature belonging to the correct bin was significantly higher when orientation was task-relevant compared to task-irrelevant during encoding. Above-chance classification for color was only present during the initial 500 ms across all conditions. Our results suggest that although information about all of an object's features is present in the initial stimulus-evoked neural response, information about the task-irrelevant features is attenuated during stimulus encoding and is largely absent throughout the delay.

Keywords

working memory; feature binding; top-down control; support-vector machine classifier

Corresponding Author Andrea Bocincova, North Dakota State University, Dept. 2765, P.O Box 6050, Fargo, ND 58108, 701.231.8934, andrea.bocincova@ndsu.edu. Jeffrey S. Johnson, jeffrey.s.johnson@ndsu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The authors declare no conflict of interest.

Footnotes

Working memory (WM) refers to our ability to temporarily store and manipulate information extracted from the ever-evolving perceptual and mental streams (Baddeley & Hitch, 1974). However, the amount of information that is available at any given moment greatly exceeds our limited storage abilities (Cowan, 2001; Luck & Vogel, 1997). To mitigate these limitations, powerful attentional selection mechanisms can be used to restrict access to WM on the basis of goal-relevant properties, including spatial location (Griffin & Nobre, 2003; Kuo, Stokes, & Nobre, 2012; Nobre, Griffin, & Rao, 2008), surface features such as color or orientation (Heuer & Schubö, 2016; Ye, Hu, Ristaniemi, Gendron, & Liu, 2016), or at the level of whole objects (Lepsien & Nobre, 2007). Less clear, however, is whether observers have voluntary control over which features of an attended multi-feature object are encoded and maintained in WM.

Studies of object-based attention have shown that when one feature of a multi-feature object is selected, attention tends to spread to other, task-irrelevant features (see, e.g., Ernst, Boynton, & Jazayeri, 2013; O'Craven, Downing, & Kanwisher, 1999). Several studies as well as prominent theoretical accounts suggest that, once selected, all features of an object are then obligatorily stored in WM regardless of their behavioral relevance (Treisman & Zhang, 2006; Vogel, Woodman, & Luck, 2001; Gao et al., 2016; Luck & Vogel, 1997). However, other evidence suggests that access to WM can be restricted to only those features that are currently relevant to the observer's goals (Serences, Ester, Vogel, & Awh, 2009; Woodman & Vogel, 2008; Yu & Shim, 2017). For example, Serences et al. (2009) used fMRI and multivariate pattern analysis (MVPA) to examine whether the individual features (color and orientation) of a single attended object can be excluded from WM when their storage is not necessary for task performance. Supporting this possibility, MVPA of BOLD responses observed during the delay period revealed that above-chance classification was only possible for the feature that was being actively maintained on that trial; the multivariate pattern of BOLD responses carried no information about the task-irrelevant feature.

Another possibility, which could potentially reconcile these views, is if all of an object's features are initially encoded, but irrelevant features can be removed from WM representations over time (for arguments to this effect, see Logie, Brockmole, & Jaswal, 2010; Marshall & Bays, 2013; Treisman & Zhang, 2006). Although the study of Serences et al. (2009) suggests that task-irrelevant object features are not held in WM throughout the delay, the use of fMRI makes it difficult to assess the time course of feature-selective processing with sufficient temporal resolution to address this possibility. In another study, Woodman and Vogel (2008) used EEG to examine differences in WM-storage related activity when all the features of a set of multi-feature objects were maintained versus only a subset. Given its high temporal resolution relative to fMRI, EEG is in principle better suited to assess the time course of feature-selective encoding and maintenance. However, their study focused on contralateral delay activity (CDA), a negative slow wave typically measured between ~400–1000 ms post stimulus that tracks the amount of information held in WM (Vogel & Machizawa, 2004). Although their findings were consistent with those of Serences et al. (2009), their focus on the CDA and delay-period activity tells us little about the time course of selective encoding and maintenance of object features in WM.

In the present study, our goal was to overcome these limitations by examining the encoding and maintenance of task-relevant and irrelevant object properties using EEG together with MVPA. In previous studies, these methods have been successfully used to assess the encoding and maintenance of both spatial and category-level information in WM with fine temporal precision (LaRocque, Lewis-Peacock, Drysdale, Oberauer, & Postle, 2012; LaRocque, Riggall, Emrich, & Postle, 2016; Lewis-Peacock, Drysdale, Oberauer, & Postle, 2011). Here, we adopted similar methods to examine the temporal evolution of neural representations reflecting the encoding and storage of task-relevant and irrelevant features of an attended two-feature object in WM. In different blocks, participants remembered the color, the orientation or both the color and the orientation of a single colored, oriented grating while the electroencephalogram was recorded from the scalp. The spatial distribution of oscillatory power across a range of frequency bands was then calculated and used to train a set of support vector machine (SVM) classifiers. First, to assess whether information about the currently relevant stimulus dimension was present in the neural signal across trials, SVM classifiers were trained to decode the task-relevant feature dimension (color vs. orientation vs. both). Secondly, to assess whether the presence of information about the identity of the object's features varied across time and depended on the feature's relevance, the SVM classifiers were trained using broad-band oscillatory activity as well as oscillatory in individual frequency bands to decode feature identity; i.e., the specific orientation and color bin the features of the remembered stimulus were drawn from.

To preview our findings, results revealed above-chance decoding of task condition (color, orientation, both) throughout each trial, including during the fixation interval preceding stimulus presentation. Additionally, the SVM classifier achieved above-chance decoding of the task-relevant feature across each condition, although, for color, this was limited to the encoding interval, with decoding accuracy falling to chance during the memory delay. Critically, above-chance decoding of task-irrelevant features was limited to the encoding interval, with classifier accuracy falling to chance beyond ~500-ms post-stimulus. Additionally, although it was possible to decode the identity of task-irrelevant features during encoding, classifier accuracy and classifier evidence, the probability of a feature belonging to the correct bin, were significantly lower than for relevant features. In summary, the results of the present study suggest differential processing and storage of features of a single object depending on their relevance to the current task.

Methods

Unless otherwise noted in the text, all reported methods, proposed analyses and specific experimental hypotheses were pre-registered through the Open Science Framework and are available together with the raw data, presentation code and analysis scripts at the following URL: osf.io/nmydf. We report how we determined our sample size, all data exclusions (if any), all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

Participants.

Twelve subjects (11 females) aged 19–31 years ($M = 23$, $SD = 3.52$) from the North Dakota State University undergraduate and graduate population participated in this study for monetary compensation (\$15/hr). All participants reported normal or corrected to normal visual acuity, normal color vision and no prior or current neurological or mental disorders. Each participant provided written informed consent and all experimental protocols were approved by the North Dakota State University IRB.

Required sample size was determined by a combination of power analysis (LaRocque et al., 2012; Samaha et al., 2016; Serences et al., 2009) and consideration of previously used sample sizes and trial numbers in studies adopting a similar approach (Foster et al., 2016; Samaha et al., 2016). A set of power analyses suggested that an N of 6 participants should be sufficient to detect significant results in each of the planned analyses. In an effort to further increase power, we doubled this number and increased the total number of trials completed per condition compared to the studies listed above.

Materials and procedure.

Stimulus presentation and response recording were controlled by a PC running Matlab (Mathworks, Inc.) with Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Stimuli were presented against a light grey background on the surface of a 19" cathode ray tube monitor with a refresh rate of 100 Hz, at a viewing distance of 60 cm.

The behavioral task consisted of a change detection working memory task in which participants were presented with a single object composed of two separate features (color and orientation), and, in separate blocks, were asked to report changes in either color (2/3 of trials), orientation (1/3 of trials) or either/both color and orientation, (1/3 of trials) between sample and test displays. The blocked design allowed for consistency of the currently task-irrelevant feature across trials within each block and so provided a high incentive for participants to only remember the task-relevant feature. Sample stimuli were identical in all experimental blocks and consisted of one centrally presented grating subtending 7.1° of visual angle. The color and orientation of the grating were drawn randomly from two distinct feature bins on each individual trial, generating four possible trial combinations of the two features (color bin 1, orientation bin 1; color bin 1, orientation bin 2; color bin 2, orientation bin 1, color bin 2, orientation bin 2)¹. Color bins were generated from a set of 180 colors equally distributed in 360° color space (CIELAB, 1976; centered at CIE $L^*a^*b^*$ coordinates: $L=70$, $A=28$, $B=12$). Color bin 1 included colors spanning the color space between 70° and 110° and color bin 2 included colors spanning the color space between 230° and 270° . Orientation bin 1 included orientations between 35° and 55° and orientation bin 2 included orientations between 125° and 145° . In order to encourage participants to remember the specific color and/or orientation of the grating, the features of the test stimuli

¹Based on the type of block, the trials were arranged into conditions as follows: *color-relevant blocks*-color-relevant condition (color bin 1, orientation bin1/bin2; color bin 2, orientation bin1/bin2) and orientation-irrelevant condition (orientation bin 1, color bin 1/bin 2; orientation bin 2, color bin1/bin 2); *orientation-relevant blocks*-orientation-relevant condition (orientation bin 1, color bin 1/bin 2; orientation bin 2, color bin 1/bin 2) and color-irrelevant condition (color bin 1, orientation bin1/bin2; color bin 2, orientation bin1/bin2); *both-relevant blocks*-color-relevant condition (color bin 1, orientation bin 1/bin 2; color bin 2, orientation bin 1/bin 2) and orientation-relevant condition (orientation bin 1, color bin 1/bin 2; orientation bin 2, color bin 1/bin 2).

were drawn from a feature space adjacent to the feature values of the sample stimuli. The amount of change required to elicit a correct response was titrated to 95% correct for each individual participant using a staircase procedure that preceded the EEG session. The titration procedure was performed separately for color and orientation to match difficulty between the features.

The sequence of trial events is shown in Figure 1. Each block began with the presentation of a cue word at the center of the screen informing the participant which feature(s) they should remember for that block (color, orientation or both color and orientation). Each trial started with the appearance of a fixation cross. After 1000 ms, a sample display was presented for 200 ms. Following an 1800-ms blank delay period, a test display was presented. On 50% of all trials, the task-relevant feature (color, orientation or either/both) was changed to a new value between sample and test. In addition to these task-relevant changes, a task-irrelevant change was introduced on 50% of trials; i.e., a change in the orientation of the grating in a color-relevant trial or a change in the color of the grating in an orientation-relevant trial. Participants were asked to ignore task-irrelevant changes and to report only whether a change in the task-relevant feature had occurred by pressing a key. This was done to ensure that participant's responses were based on detecting changes only along the task-relevant dimension. Although responses were not speeded, the time to respond was limited to 2 s. The experiment consisted of 1,728 total trials (576 per feature dimension and 288 per feature bin) separated into two sessions with 18 blocks. Blocks with different experimental conditions were randomly intermixed across each session and trials within a block were presented in random order. Participants completed 10 practice trials from each block type prior to the experiment. The experiment took ~2 h per session to complete.

EEG data acquisition and preprocessing.

EEG was recorded from 64 scalp sites using active Ag/AgCl electrodes (BioSemi Active Two) positioned according to the standards of the modified international 10/20 system (American Electroencephalographic Society, 1994) and two additional electrodes positioned over the left and right mastoids. The common mode sense (CMS) electrode was located at site C1, with a driven right leg (DRL) electrode located at site C2. To monitor blinks and eye movements, the electrooculogram (EOG) was recorded from electrodes placed above and below each eye and ~1cm to the left and right of the external canthi of each eye. EEG and EOG were recorded with a sampling rate of 512 Hz and a band-pass filter of 0.01–100 Hz.

Data were processed off-line using the EEGLab open source Matlab-based toolbox (Delorme & Makeig, 2004) and custom-written analysis scripts in Matlab (MathWorks, Inc., Natic, MA, USA). All signals were re-referenced offline to the algebraic average of the left and right mastoids, downsampled to 256 Hz, and epoched into 4000 ms long segments spanning the time interval 1000 ms before to 3000 ms after sample display onset. To detect eye movements and blinks in the epoched data, EOG signals were combined to derive bipolar vertical and horizontal channels that were passed through a set of artifact detection steps. To avoid any contamination of data by eye blinks or muscle movements, any trials containing amplitude changes larger than 80 μ V in the vertical bipolar EOG channel and 120 μ V in any other channel within a moving window of 200 ms were removed. Additionally, to

prevent contamination due to eye movements, a step function was used to detect amplitude changes larger than 25 μV in the horizontal bipolar EOG channel. Participants were removed from further analysis if the number of trials rejected exceeded 30% of total trials (cut-off established prior to the beginning of the study). Rejection rate was under 30% of total trials for all participants ($M = 7.01$, $SD = 2.88$).

Data Analysis

Behavioral performance.—The proportion of correct trials was calculated for each condition (color-relevant, orientation-relevant, both-relevant) as the number of correct trials divided by the total number of trials. Statistical significance of differences in proportion correct between experimental conditions was evaluated using one-way ANOVA. This analysis was performed to ensure that any differences in classifier accuracy were not caused by significant differences in participants' performance across experimental conditions.

EEG time-frequency analysis.—All raw EEG data were Laplace transformed prior to time-frequency analysis to reduce volume conduction and maximize the spatial resolution of the signal (Carvalhaes & de Barros, 2015). To extract instantaneous oscillatory power values, the Laplace transformed signal was first bandpass filtered into a range of frequency bands between 4 and 34 Hz in steps of 1 Hz using a two-way least squares finite impulse response filter using the EEGLab toolbox function `eegfilt.m` (see e.g., Foster et al., 2016 for the same approach). This function uses a zero-phase forward and reverse operation that eliminates phase shifting caused by the use of forward-only filters. The filtered data, $f(t)$, were then Hilbert transformed using functions included in the MATLAB Signal Processing Toolbox to produce the complex analytic signal, $A(t)e^{i\phi(t)}$, where $A(t)$ is the instantaneous amplitude and $\phi(t)$ is the instantaneous phase. Frequency specific instantaneous amplitudes were then extracted from the complex analytical signal using the following MATLAB syntax:

$$\text{hilbert}\left(\text{eegfilt}\left(\text{data}, F_{\text{sample}}, f_1, f_2\right)\right)'$$

where *data* is a 2D matrix of EEG data (number of trials x number of time points), F_{sample} is the sampling frequency (256 Hz) and f_1 and f_2 are the lower and higher bounds of the frequency band of interest, respectively.

Calculating evoked and total power.—The instantaneous amplitudes were then used to calculate separate estimates of single-trial evoked and total (induced) oscillatory power. To do so, trials from each experimental condition were randomly assigned to 10 separate blocks. The number of trials was kept equal across blocks and experimental conditions and was calculated by dividing the total number of trials in the condition with the lowest number of trials by the number of blocks, and rounding the number down to the nearest integer. Importantly, to prevent circularity, no trial was ever assigned to multiple blocks. Additionally, to avoid any idiosyncrasies generated by the assignment of specific trials to blocks and the possibility that a subset of trials might be omitted in conditions with higher

numbers of trials, the process of randomized assignment of trials into blocks and the following classification was repeated 10 times.

Evoked power was calculated by first averaging the complex signal across trials within each block, and then squaring the absolute value of the averaged signal. Evoked power values calculated in this way reflect activity that is phase-locked to the time-locking event (i.e., the memory display), because averaging prior to squaring cancels out signals that are out of phase. By contrast, total power was extracted by squaring the absolute value of the complex analytic signal separately for each trial and then averaging across all trials within each block. Total power therefore reflects activity that is time-locked to stimulus onset as well as activity that may be out of phase.

Multivariate pattern analysis.—The spatial distribution of oscillatory power across the scalp was used to train a set of SVM classifiers using LIBSVM software (Chang & Lin, 2011) running under MATLAB. First, to assess whether the spatial distribution of EEG activity carried information about which feature(s) of the object is currently relevant, the SVM classifiers were trained to decode the task-relevant feature dimension (color vs. orientation vs. both). Secondly, to assess whether the presence of information about the identity of the object's features varied across time and depended on the feature's relevance, the SVM classifiers were trained to decode the specific orientation and color bin (bin 1 or bin 2) the features of the remembered stimulus were drawn from. Data from experimental conditions that were to be classified (e.g., color bin 1 versus color bin 2) were assigned a category label (e.g., -1 and 1) and were concatenated. Data from all blocks but one (the “training” set) were used to train the classifier. The trained classifier was then used to classify the remaining block (the “test” set). Using this hold-one-out procedure, the classification was cross-validated until data from every block served as the test set. Depending on whether the assigned label corresponded to the actual exemplar label, the output of the classifier was marked as correct (1) or incorrect (0). Accuracy values were averaged across the cross-validation results and the 10 iterations to generate a measure of *classifier accuracy*.

In order to track changes in classification accuracy over time, the classification was performed separately for each time point. The classification was done in three rounds. In the *first round*, classification was performed using the power values from all frequencies at once (each block containing values from 64 channels x 31 frequencies = 1,984 unique features) resulting in one classifier accuracy value for each time point (i.e. time-series classifier accuracy). In the *second round*, classification was performed for one frequency at a time (each block containing power values from 64 channels, i.e., 64 unique features used for classification), which generated 31 frequency-specific classifier accuracy values for each time point (i.e., time-frequency series classification)².

²We also assessed whether the maintenance of task-relevant or irrelevant information in WM was supported by oscillatory activity within specific frequency bands of the EEG, as has been observed in other studies of storage in WM (see, e.g., Foster et al., 2016; Foster, Sutterer, Serences, Vogel, & Awh, 2017; Samaha et al., 2016). In a separate classification run (see Supplementary materials for results), classification was performed across three specific frequency bands (theta 3–8Hz, alpha 8–12Hz, beta 15–30Hz). In this case, power values were averaged across frequencies belonging to a specific frequency band prior to classification (each block containing averaged frequency band power values for 64 channels, i.e., 64 unique features used for classification) resulting in 3 frequency-band-specific classifier accuracy values for each time point.

Statistical significance of the observed classifier accuracy was evaluated against a null distribution using one-sample t-tests. A null distribution was generated by randomly reassigning labels to the classifier training sets, thus eliminating any systematic relationship between the experimental condition and the spatial pattern of power values, and running the classification procedure as described above. This randomization procedure was repeated 1,000 times, resulting in a null distribution of t-statistics for each frequency and time-point (or time-point only for classification across frequencies). To increase computational speed, data were downsampled to 50 Hz prior to the randomization procedure.

In addition to the categorical output, we were also interested in the probability of the test block belonging to each of the classified categories, or in other words, the classifier's degree of certainty about the assigned label. Because SVM classifiers do not automatically provide posterior probabilities in their output, probabilities were calculated by running a separate SVM classifier with the same settings used for establishing classifier accuracy with the exception that Platt scaling (LIBSVM; Chang & Lin, 2011) was used to derive probability distributions over classes (Platt, 1999). This probability, here referred to as *classifier evidence*, was used to compare the amount of evidence available in the neural signal over time for the different categories (bin 1 versus bin 2) in each block type (orientation, color, both). Classifier evidence was compared between conditions (relevant, irrelevant, both) separately for orientation and color for all individual time points. Benjamini-Hochberg correction for multiple comparisons was applied to the resulting p-values. Note that this analysis diverged from the originally pre-registered analysis that proposed to compare classifier evidence for bin-present versus bin-absent trials; i.e., classifier evidence for a particular bin (e.g., color bin 1) calculated for trials in which the stimulus contained this feature bin (present; e.g., color bin 1 orientation bin 2) and for trials in which the stimulus contained the other feature bin (absent; e.g., color bin 2 orientation bin 2). Because classifier evidence for bin present and bin absent trials always sums to $P = 1$, and because $P = 0.5$ denotes equal classifier evidence for both bins, their values mirror each other (e.g., if classifier evidence for bin present is 0.64, classifier evidence for bin absent is $1 - 0.64$, or 0.36, which corresponds to 0.14 above and below $P = 0.5$ respectively). Clearly, this planned comparison is not as informative as comparing classifier evidence for only bin present trials between the different experimental conditions. The results of the originally proposed analysis can be found in the Supplemental Materials.

Results

Behavioral performance.

Analysis of proportion correct recognition revealed uniformly high performance across the three experimental conditions [color ($M = 89.66$, $SD = 1.20$), orientation ($M = 91.40$, $SD = 1.19$), both ($M = 89.38$, $SD = 0.92$)], $F(2,22) = 1.822$, $p = .185$, $\eta_p^2 = .142$. Proportion correct increased overall from session one to session two, $F(1,11) = 5.218$, $p = .043$, $\eta_p^2 = .322$, possibly reflecting an effect of practice with the task. Overall, these results indicate that the titration procedure successfully equated difficulty between the three experimental conditions.

Decoding task-relevant feature dimension.

To test whether the spatial distribution of EEG activity contained information about the currently relevant feature dimension, the classifier was trained to differentiate between the three experimental conditions (color-relevant, orientation-relevant, both-relevant). The results suggest that above-chance classification of experimental condition was possible throughout the entire trial, including during the pre-stimulus interval, during encoding and throughout the delay (Figure 2A). Above chance classifier accuracy prior to the presentation of the stimulus could represent preparatory activity, because the feature that was task-relevant was manipulated at the level of experimental blocks. It is also possible that the use of a blocked design allowed the classifier to rely on more general changes in power happening at the level of experimental blocks (reflecting, e.g., cognitive set, noise, or other sources of variance) rather than relying on trial-specific activity reflecting storage of the task-relevant feature(s).

Interestingly, repeated measures ANOVA comparing classifier accuracy between the pre-stimulus (−200–0 ms), encoding (0–300 ms) and delay (300–2000 ms) intervals showed significant differences in decoding accuracy, $F(2,22) = 9.581$, $p = .001$, $\eta_p^2 = .466$ (Figure 2B). Follow-up pairwise comparisons showed no significant difference in decoding accuracy between the encoding and delay period ($p = .893$); however, there was a significant increase in decoding accuracy between the pre-stimulus interval and both the encoding and delay periods ($ps < .006$). This suggests that the patterns of oscillatory power following stimulus-presentation carried additional information about the task-relevant feature dimension that was not present during the pre-stimulus interval.

As shown in Figure 2C, above-chance decoding during the pre-stimulus interval was supported by activity in the high beta band (~20–34 Hz). By contrast, decoding during stimulus encoding (~−100–400 ms) was supported by oscillatory power in the theta (4–8 Hz) and beta bands, whereas decoding during the delay period (~200–2000 ms) was supported by increased power in the alpha (8–15 Hz) and higher frequency ranges.

To assess whether information about individual features varied across time and as a function of feature relevance, SVM classifiers were trained to decode the orientation and color of the memory stimulus; i.e., which orientation and color bin the item's features were drawn from. If all features of a multi-feature object are attended and automatically encoded into WM irrespective of their relevance, the identity of both task-relevant and task-irrelevant features should be decodable throughout both the encoding and delay periods. However, if access to WM can be restricted to only task-relevant information, only the identity of the task-relevant feature should be decodable throughout the delay.

Decoding feature identity from stimulus-evoked neural activity.

As can be seen in the separate panels of Figure 3, both the relevant and irrelevant features of an attended object were decodable from the distributed pattern of EEG activations evoked by the stimulus (i.e., from signals time-locked to stimulus onset). Significant decoding persisted for up to ~700 ms irrespective of feature relevance and for both color and orientation stimuli.

The following two sections describe in greater detail the decoding results for the separate dimensions of orientation and color.

Orientation.—As shown in Figure 3A, analysis of across-frequency, time-series classifier accuracy showed that significant above-chance classifier performance lasted the longest in the orientation-relevant condition (~1200 ms) followed by the both-relevant condition (~800 ms) and the orientation-irrelevant condition (~700 ms). As can be seen, classifier accuracy rises quickly in the interval shortly before the presentation of the sample stimulus and peaks between ~100–400 ms post-stimulus before declining back to baseline levels. (Note that above-chance decoding prior to stimulus onset likely reflects temporal smearing of EEG power due to time-frequency decomposition).

A 3 (feature relevance: relevant, irrelevant, both) x 3 (time interval: pre-stimulus, encoding and delay) repeated measures ANOVA comparing average decoding accuracy across a-priori defined time intervals revealed a main effect of feature, $F(2,22) = 10.168$, $p = .001$, $\eta_p^2 = .480$ (Figure 3B). Follow-up paired comparisons revealed significantly higher classifier accuracy for feature relevant ($M = .705$, $SD = .019$) versus irrelevant ($p = .003$; $M = .653$, $SD = .020$) or both-relevant ($p = .001$; $M = .657$, $SD = .020$) conditions. There was no difference in classifier accuracy between the irrelevant and both-relevant ($p = .765$) conditions. There was also a main effect of time interval, $F(2,22) = 88.893$, $p < .001$, $\eta_p^2 = .890$, suggesting higher decoding accuracy during encoding ($M = .872$, $SD = .022$) compared to both pre-stimulus ($p < .001$; $M = .582$, $SD = .033$) and delay ($p < .001$; $M = .562$, $SD = .009$) intervals. There was no difference in decoding accuracy between pre-stimulus and delay intervals ($p = .568$). There was also no significant interaction of feature relevance and time interval $F(2,22) = 1.998$, $p = .110$, $\eta_p^2 = .154$.

Time-frequency decoding analysis (Figure 3C) revealed that, for all conditions (relevant, irrelevant, both), above-chance decoding of orientation during the initial ~500 ms of each trial was supported by evoked oscillatory power spanning a range of frequencies (~4–30 Hz).

Color.—As with orientation, across-frequency time-series decoding for color showed that decoding accuracy began increasing around 200 ms prior to stimulus presentation and peaked ~100–500 ms later before falling back to chance levels ~800 ms later (Figure 3D). Above chance decoding persisted the longest when color was task-relevant (~1000 ms) compared to when it was irrelevant (~850 ms) or stored together with orientation (~750 ms).

Repeated measures ANOVA (Figure 3E) revealed a significant main effect of time interval, $F(2,22) = 84.377$, $p < .001$, $\eta_p^2 = .885$, reflecting higher classification accuracy during encoding ($M = .855$, $SD = .023$) compared to both pre-stimulus ($p < .001$; $M = .593$, $SD = .030$) and delay ($p < .001$; $M = .560$, $SD = .006$). There was no difference in decoding accuracy between the pre-stimulus and delay intervals ($p = .305$). The main effect of feature relevance and the interaction between time interval and feature relevance were not significant [$F(2,22) = 2.756$, $p = .085$, $\eta_p^2 = .200$ and $F(2,22) = 0.295$, $p = .880$, $\eta_p^2 = .026$ respectively].

As shown in Figure 3F, and in keeping with the analysis of orientation trials, time-frequency decoding analysis showed that significant above-chance decoding of color during the initial ~500 ms following stimulus onset was supported by oscillatory power across a broad range of frequencies (~4–30 Hz).

Decoding feature identity from total power.

To test whether WM storage was restricted to the task-relevant features of an attended object, we performed a second decoding analysis focusing on decoding feature identity from EEG activity induced by the stimulus (i.e., signals not strictly time-locked to stimulus onset).

Orientation.—In line with our predictions, decoding of the bin that the orientation was drawn from was consistently above chance throughout the whole delay period when orientation was task relevant. When orientation was irrelevant, above-chance decoding only persisted for ~600 ms following stimulus onset. Importantly, when both orientation and color needed to be remembered, orientation decoding accuracy remained above chance across the full delay period except for two small ~200 and 250 ms interruptions around 1200 and 1700 ms post-stimulus (Figure 4A).

To evaluate differences in classifier accuracy between the three time intervals based on feature relevance, we performed a 3 (time interval: pre-stimulus, encoding, delay interval) x 3 (feature: relevant, irrelevant, both) repeated measures ANOVA (Figure 4B). The ANOVA revealed a main effect of feature, $F(2,22) = 7.248$, $p = .004$, $\eta_p^2 = .397$. Follow-up pairwise comparisons revealed lower decoding accuracy in the task-irrelevant ($M = 0.513$, $SD = 0.016$) versus task-relevant ($M = 0.616$, $SD = 0.028$; $p = .008$) and both-relevant ($M = 0.570$, $SD = 0.016$; $p = .019$) conditions. There was, however, no significant difference in decoding accuracy when orientation was the only task-relevant feature or both color and orientation were task-relevant ($p = .125$). Additionally, there was a main effect of time interval, $F(2,22) = 18.383$, $p < .001$, $\eta_p^2 = .626$. Follow-up pairwise comparisons revealed a significant increase in decoding accuracy between the pre-stimulus ($M = 0.471$, $SD = 0.016$) and the encoding ($M = .631$, $SD = 0.025$; $p < .001$) and delay intervals ($M = .597$, $SD = 0.021$; $p < .001$). The difference in decoding accuracy between the encoding and delay intervals was not significant ($p = .334$).

Unlike the analyses of decoding based on evoked power, there was a significant interaction between feature relevance and time interval for total power, $F(4,44) = 6.262$, $p = .003$, $\eta_p^2 = .363$. Follow-up comparisons for the significant interaction showed that during the pre-stimulus interval there was no significant difference in decoding accuracy between conditions when orientation was task-relevant ($M = 0.456$, $SD = 0.036$) or task-irrelevant ($M = 0.445$, $SD = 0.023$; $p = .814$). There was also no difference between task-relevant and both-relevant conditions ($M = 0.511$, $SD = 0.023$; $p = .227$) during the pre-stimulus interval. There was a small statistical difference between the task-irrelevant and both-relevant conditions ($p = .014$, Bonferroni corrected to p -val of $p < .016$). However, one-sample t-tests comparing observed decoding accuracy to both theoretical and empirical (i.e., derived from null distributions) chance showed that decoding accuracy did not exceed chance in any of

the three conditions (all p s $> .034$, Bonferroni corrected to significant p -val of $p = .006$). This suggests that during the pre-stimulus interval, decoding accuracy was below chance levels irrespective of feature relevance.

More importantly, during the encoding interval, there was no statistically significant difference in decoding accuracy between task-irrelevant ($M = 0.574$, $SD = 0.029$) and both-relevant conditions ($M = 0.617$, $SD = 0.024$; $p = .292$). Decoding accuracy when orientation was task-relevant ($M = 0.702$, $SD = 0.046$) was higher than when it was task-irrelevant ($p = .014$ Bonferroni corrected) but not significantly different from when both color and orientation were task relevant ($p = .063$). A one sample t-test against both theoretical and empirical chance showed that decoding accuracy was significantly above chance for all three conditions during encoding (all p s $< .001$, Bonferroni corrected to significant p -val of $p = .006$). These results suggest that, during the encoding period, decoding of orientation identity was above chance irrespective of task relevance, although decoding accuracy was higher when orientation was task relevant.

Finally, during the delay interval, decoding accuracy was significantly higher when orientation was task-relevant ($M = 0.690$, $SD = 0.029$) compared to when orientation was task-irrelevant ($M = 0.521$, $SD = 0.029$; $p < .001$ Bonferroni corrected) or when both orientation and color were task-relevant ($M = 0.581$, $SD = 0.024$; $p < .001$, Bonferroni corrected). Importantly, decoding accuracy when both orientation and color were task-relevant was significantly higher than decoding accuracy in the task-irrelevant condition ($p = .012$ Bonferroni corrected). One-sample t-tests against both theoretical and empirical chance revealed statistically significant decoding accuracy for orientation-relevant (p s $< .001$) and both-relevant conditions (theoretical chance $p = .006$; empirical chance $p < .001$, Bonferroni corrected to significant p -val of $p = .006$). However, decoding accuracy for the orientation-irrelevant condition did not exceed theoretical ($p = .326$, Bonferroni corrected to significant p -val of $p = .006$) or empirical chance ($p = .012$, Bonferroni corrected to significant p -val of $p = .006$), suggesting that orientation identity was not decodable during the delay period when orientation was task-irrelevant. Overall, these results provide support for the possibility that information about feature identity is initially present in all conditions (albeit to different degrees), but persists only if the storage of orientation is required by the task (i.e., orientation-relevant, both-relevant).

In keeping with the broad-band analysis, decoding on the basis of the full matrix of time-frequency values across electrodes revealed above chance decoding throughout the delay, but only when orientation was the task-relevant feature (Figure 4C). Observed above chance decoding during the initial ~500 ms was supported by oscillatory activity in the theta band (~4–8 Hz) in all three conditions. By contrast, when orientation was the task-relevant feature or was remembered together with color, above-chance decoding during the maintenance period (~500 – 2000 ms) was supported by power in the alpha band (~8–14 Hz).

Color.—Unlike classification of orientation, classification of color using induced power values did not produce consistent above-chance decoding accuracy across the maintenance period for any of the experimental conditions examined (color-relevant, color-irrelevant, both-relevant). Time-series decoding accuracy analysis showed that color identity was only

decodable at above chance levels during the first ~800–1000ms of each trial (Figure 4D). As observed for orientation, classification based on the matrix of time-frequency values across electrodes revealed that above chance decoding during the initial ~500 ms was supported by oscillatory activity in the theta-band (Figure 4F).

As with orientation, a 3×3 repeated measures ANOVA (Figure 4E) was performed to test differences in classifier accuracy between the three time intervals (pre-stimulus, encoding, delay) based on the relevance of the feature (relevant, irrelevant, both-relevant). ANOVA only revealed a main effect of time interval $F(2,22) = 9.748$, $p = .001$, $\eta_p^2 = .470$. Follow-up pairwise comparisons revealed that decoding accuracy was higher across all conditions during the encoding interval ($M = 0.597$, $SD = 0.032$) compared to both pre-stimulus ($M = 0.467$, $SD = 0.020$; $p = .004$) and delay ($M = 0.505$, $SD = 0.010$; $p = .017$, not significant after Bonferroni corrected to p-val of $p < .016$) intervals. There was no significant difference in decoding accuracy between the pre-stimulus and delay period intervals ($p = 0.83$). There was also no main effect of feature, $F(1.377, 15.145) = 2.901$, $p = .100$, $\eta_p^2 = .209$ (Greenhouse-Geisser corrected), and no significant interaction between feature and time interval $F(4,44) = 1.517$, $p = .214$, $\eta_p^2 = .121$.

Classifier Evidence.

Results presented thus far suggest that feature decoding accuracy was uniformly high during stimulus processing and encoding, irrespective of whether the feature being decoded was task-relevant, irrelevant, or remembered together with a feature from the other dimension. Conversely, above-chance decoding throughout the full delay interval was only possible when the decoded feature was task-relevant, although this was less reliable when it was remembered together with color. However, these findings were based on an analysis of classifier accuracy, which tells us whether the decision value generated by the classifier corresponded to the correct category, but gives us no information about the relative differences in likelihood for each option. In order to further evaluate the differences in decoding of task-relevant and task-irrelevant features, therefore, we conducted an additional analysis of classifier evidence. Classifier evidence reflects the posterior probability, or the likelihood, that a given pattern of neural activity corresponds to a given trial type (e.g., orientation bin 1 vs. orientation bin 2). As such, it is a more sensitive measure of neural processing than classifier accuracy because it does not require a forced-choice selection of the most likely of a range of options (e.g., whether orientation 1 or orientation 2 was remembered).

To evaluate differences in classifier evidence between conditions, a series of paired t-tests were performed. Classifier evidence derived from both evoked and total power was significantly higher in orientation-relevant versus orientation-irrelevant trials throughout the initial ~700 ms of the trial, which includes the encoding interval ($p < .05$, Benjamini-Hochberg corrected; Figure 5AB). This means that even though the classifier could decode orientation at above-chance levels during encoding whether orientation was relevant or irrelevant, there was substantially more evidence in the neural signal about which bin the orientation was drawn from when orientation was task-relevant. There was also a significant difference in evoked power classifier evidence between orientation relevant and both

relevant conditions in the early trial period. There were no significant differences in classifier evidence across conditions for color ($ps > .05$, Benjamini-Hochberg corrected; Figure 5CD).

Discussion

The goal of the current study was to assess how the representation of individual features of an attended multi-feature object in WM varies over time depending on the features' relevance to the ongoing task. To do this, we tracked the temporal evolution of neural representations reflecting the task-relevant and irrelevant features in WM using SVM classifiers trained on the spatial distribution of oscillatory power over the time course of an experimental trial. To assess whether the pattern of oscillatory power changed depending on what features of a multi-feature object were task-relevant, we tested whether SVM classifiers could be successfully trained to correctly decode the currently relevant feature dimension (color, orientation, both). To assess whether the spatial patterns of oscillatory power represented the identity of the object's features and whether this changed over time depending on the feature's relevance, we tested whether SVM classifiers could accurately decode which bin the object's features were drawn from.

The results showed differential processing and storage of features of a single, multi-feature object depending on their relevance to the current task. First of all, variations in oscillatory power supported above-chance decoding of the task-relevant feature dimension (whether color, orientation, or both were to be remembered for a given block of trials). Interestingly, the relevant feature dimension was decodable throughout the entire trial, including during the fixation interval prior to the appearance of the memory stimulus. This suggests that task-level information about which feature was relevant was present in the neural signal throughout each block of trials and was not dependent on the presentation of the memory stimulus on each trial. Secondly, our findings revealed that successful decoding of feature identity varied over time depending on the feature's relevance in the following way. During stimulus processing and encoding, the specific features that were seen on a given trial could be decoded from patterns of evoked power irrespective of their relevance. Conversely, analysis of total power revealed differences in the ability to decode feature identity depending on feature relevance. For orientation, feature identity was decodable during the early stages of a trial irrespective of feature relevance, but only task-relevant orientations were decodable across the full delay; above chance decoding of orientation in the task-irrelevant condition was limited to the initial ~500 ms of the trial.

Interestingly, time-frequency-based decoding revealed that the identity of task-relevant orientations throughout the delay was carried by neural activity in the alpha frequency band (~8–14 Hz; see Figure S1 and Supplemental Materials for additional analyses and discussion of band-specific decoding performance). Previous studies have suggested that activity in the alpha frequency band carries information about the spatial location of remembered features (Foster, Bsaies, Jaffe, & Awh, 2017; Foster et al., 2016). In keeping with this finding, Bae and Luck (2017) showed that alpha-band activity tracks stimulus location (i.e., environment-centered spatial representations), whereas information about stimulus orientation (i.e., object-centered spatial representations) is carried by sustained, stimulus-specific ERPs. In

light of these findings, it is possible that the orientation decoding results described here could reflect something other than orientation coding per se; for example, stimulus orientation could be remembered by focusing attention on a particular location throughout the delay (e.g., somewhere to the left for left-tilting stimuli, or to the right for right-tilting stimuli). Although the present study was not designed to differentiate between these possibilities, it is important to note that the classifier's performance was sensitive to feature relevance, which suggests that alpha-band activity carried information relevant to the remembered feature.

The analysis of classifier accuracy summarized above suggests that orientation decoding was high across conditions throughout the early encoding interval. However, a direct comparison of classifier evidence for each of the conditions suggested that the probability of correctly classifying the feature based on the neural signal varied substantially across conditions. In particular, when the to-be-decoded orientation was task-irrelevant, classifier evidence in favor of the correct feature derived from total power was only slightly higher than evidence for the incorrect feature, and only remained above chance for ~500 ms following stimulus onset. By contrast, when orientation was task relevant, classifier evidence strongly supported the correct feature during encoding and persisted throughout the full delay interval. Significant differences in classifier evidence between task-relevant and irrelevant conditions were also found in evoked power. This suggests that, although accurate decoding was possible irrespective of feature relevance, neural activity reflecting the processing of task-irrelevant information was greatly attenuated even during initial stimulus encoding.

In contrast to the results for orientation, above-chance decoding of color was limited to the stimulus processing and encoding interval and was not selective for feature relevance. Our choice to use color as one of the feature dimensions was based on the results of a pilot study ($n=3$) that led us to believe that successful decoding of color from the EEG would be possible, so we were somewhat surprised by these results. Despite the lack of above-chance decoding, participants were able to successfully remember the color of the object as evidenced by well above-chance behavioral performance. One possible explanation for the lack of successful color decoding observed here is that, unlike orientation, color information is maintained in brain areas with anatomical characteristics that do not generate oscillatory signals that are traceable by scalp-recorded EEG. Similar difficulty in decoding color information across the maintenance period of a WM task has previously been reported in a study using local field potentials in monkeys (Lara & Wallis, 2014). Based strictly on our results, it is unclear whether color can be selectively excluded from a WM representation when task-irrelevant.

The results of the current study are relevant to our understanding of the neural fate of task-relevant and irrelevant object features. In line with previous findings, the results showed that task-irrelevant features were not actively stored in WM across the maintenance period (Serences et al., 2009; Woodman & Vogel, 2008). Extending these findings, results of the current study suggest that the representation of task-irrelevant features is already greatly attenuated at the time of the object's initial processing and encoding. This finding appears at odds with the proposal that all of an attended object's features are initially encoded, with task-irrelevant features being progressively removed from the WM representation over time

(Logie et al., 2010). Instead, our observations demonstrate that although information regarding all of an object's features is present in the initial stimulus-evoked neural response, information about task-irrelevant features is greatly attenuated during stimulus encoding and is entirely absent throughout the delay interval.

Even though our results are consistent with the selective storage of task-relevant features in WM, another possible explanation of our findings is that all features of an object are encoded and stored in WM, but are maintained differently depending on their relevance to the task. Prominent theoretical accounts of WM have proposed that different objects stored in WM can be held in different states of activation depending on the current focus of attention (Cowan, 1995; Oberauer, 2002). In keeping with this proposal, a number of studies have shown that when an object is currently unattended, but is still maintained in WM, its representation is not decodable from the neural signal until attention is redirected back to its representation (LaRocque et al., 2012, 2016; Lewis-Peacock et al., 2011; Rose et al., 2016). These results have been taken as support for the proposal that unattended items in WM are stored in a so-called "silent state" characterized by subthreshold patterns of neural activation (Mongillo, Barak, & Tsodyks, 2008; Stokes, 2015; Stokes et al., 2013). Future studies should examine whether similar mechanisms operate at the level of individual object features.

Another possibility is that, rather than reflecting attenuated coding of irrelevant features across trials, the ability of participants to filter out the task-irrelevant feature during encoding varied across trials. In this scenario, the level of decoding accuracy achieved for task-irrelevant features would represent a mixture of trials on which the task-irrelevant feature was completely filtered out and trials on which it was not. Due to the nature of our classification procedure, in which trials were averaged into blocks to increase the signal-to-noise ratio, we were not able to analyze whether decoding accuracy for the task-irrelevant feature on individual trials was associated with behavioral performance (e.g., whether higher classifier accuracy for task-irrelevant features increased the likelihood of a false alarm in a task-irrelevant change trial). Additionally, due to a relatively small sample size and relatively low variance in behavioral performance, we were not able to evaluate the relationship between decoding of the task-irrelevant feature and behavioral performance. Future research should focus on addressing this relationship.

In summary, the results of the current study suggest that task-relevant features of an attended multi-feature object can be selectively encoded and maintained in WM. It appears that the representation of task-irrelevant features is significantly weaker compared to task-relevant features during the earliest stages of processing and encoding. These results add to the converging evidence suggesting that in contrast to the object-based deployment of attention during perception, a priority can be given to task-relevant features when object information is encoded and maintained in WM.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the National Institute of Health under Grant R15 MH105866-03 awarded to Jeffrey S. Johnson and Grant P30 GM103505-02 awarded to Mark McCourt.

References

- Baddeley AD, & Hitch GJ (1974). Working Memory. In Bower GH (Ed.), *The psychology of Learning and Motivation* (pp. 47–90). New York: Academic Press.
- Bae G-Y, & Luck SJ (2017). Dissociable Decoding of Spatial Attention and Working Memory from EEG Oscillations and Sustained Potentials. *Journal of Neuroscience*, 2860–17. 10.1523/JNEUROSCI.2860-17.2017
- Brainard DH (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436. [PubMed: 9176952]
- Carvalhoes C, & de Barros JA (2015). The surface Laplacian technique in EEG: Theory and methods. *International Journal of Psychophysiology*, 97(3), 174–188. 10.1016/j.ijpsycho.2015.04.023 [PubMed: 25962714]
- Chang C-C, & Lin C-J (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 1–27. 10.1145/1961189.1961199
- Cowan N (1995). *Attention and memory: An integrated framework* New York: Oxford University Press.
- Cowan N (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185. [PubMed: 11515286]
- Delorme A, & Makeig S (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134, 9–21. [PubMed: 15102499]
- Ernst ZR, Boynton GM, & Jazayeri M (2013). The spread of attention across features of a surface. *Journal of Neurophysiology*, 110(10), 2426–2439. 10.1152/jn.00828.2012 [PubMed: 23883860]
- Foster JJ, Bsaies EM, Jaffe RJ, & Awh E (2017). Alpha-Band Activity Reveals Spontaneous Representations of Spatial Position in Visual Working Memory. *Current Biology*, 27(20), 3216–3223.e6 10.1016/j.cub.2017.09.031 [PubMed: 29033335]
- Foster JJ, Sutterer DW, Serences JT, Vogel EK, & Awh E (2016). The topography of alpha-band activity tracks the content of spatial working memory. *Journal of Neurophysiology*, 115(1), 168–177. 10.1152/jn.00860.2015 [PubMed: 26467522]
- Foster JJ, Sutterer DW, Serences JT, Vogel EK, & Awh E (2017). Alpha-Band Oscillations Enable Spatially and Temporally Resolved Tracking of Covert Spatial Attention. *Psychological Science*, 0956797617699167 10.1177/0956797617699167
- Gao Z, Yu S, Zhu C, Shui R, Weng X, Li P, & Shen M (2016). Object-based Encoding in Visual Working Memory: Evidence from Memory-driven Attentional Capture. *Scientific Reports*, 6 10.1038/srep22822
- Griffin IC, & Nobre AC (2003). Orienting Attention to Locations in Internal Representations. *Journal of Cognitive Neuroscience*, 15(8), 1176–1194. 10.1162/089892903322598139 [PubMed: 14709235]
- Heuer A, & Schubö A (2016). Feature-based and spatial attentional selection in visual working memory. *Memory & Cognition*, 44(4), 621–632. 10.3758/s13421-015-0584-5 [PubMed: 26754949]
- Kuo B-C, Stokes MG, & Nobre AC (2012). Attention modulates maintenance of representations in visual short-term memory. *Journal of Cognitive Neuroscience*, 24(1), 51–60. 10.1162/jocn_a_00087 [PubMed: 21736457]
- Lara AH, & Wallis JD (2014). Executive control processes underlying multi-item working memory. *Nature Neuroscience*, 17(6), 876–883. 10.1038/nn.3702 [PubMed: 24747574]
- LaRocque JJ, Lewis-Peacock JA, Drysdale AT, Oberauer K, & Postle BR (2012). Decoding Attended Information in Short-term Memory: An EEG Study. *Journal of Cognitive Neuroscience*, 25(1), 127–142. 10.1162/jocn_a_00305

- LaRocque JJ, Riggall AC, Emrich SM, & Postle BR (2016). Within-Category Decoding of Information in Different Attentional States in Short-Term Memory. *Cerebral Cortex* (New York, N.Y.: 1991) 10.1093/cercor/bhw283
- Lepsien J, & Nobre AC (2007). Attentional Modulation of Object Representations in Working Memory. *Cerebral Cortex*, 17(9), 2072–2083. 10.1093/cercor/bhl116 [PubMed: 17099066]
- Lewis-Peacock JA, Drysdale AT, Oberauer K, & Postle BR (2011). Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. *Journal of Cognitive Neuroscience*, 24(1), 61–79. 10.1162/jocn_a_00140 [PubMed: 21955164]
- Logie RH, Brockmole JR, & Jaswal S (2010). Feature binding in visual short-term memory is unaffected by task-irrelevant changes of location, shape, and color. *Memory & Cognition*, 39(1), 24–36. 10.3758/s13421-010-0001-z
- Luck SJ, & Kappenman ES (2012). *The Oxford Handbook of Event-Related Potential Components* Oxford University Press, USA.
- Luck SJ, & Vogel EK (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281. 10.1038/36846 [PubMed: 9384378]
- Marshall L, & Bays PM (2013). Obligatory encoding of task-irrelevant features depletes working memory resources. *Journal of Vision*, 13(2). 10.1167/13.2.21
- Mongillo G, Barak O, & Tsodyks M (2008). Synaptic Theory of Working Memory. *Science*, 319(5869), 1543–1546. 10.1126/science.1150769 [PubMed: 18339943]
- Nobre AC, Griffin IC, & Rao A (2008). Spatial Attention can Bias Search in Visual Short-Term Memory. *Frontiers in Human Neuroscience*, 1 10.3389/neuro.09.004.2007
- Oberauer K (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 411–421. 10.1037/0278-7393.28.3.411
- O’Craven KM, Downing P, & Kanwisher N (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, 401(6753), 584–587. 10.1038/44134 [PubMed: 10524624]
- Pelli DG (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442. [PubMed: 9176953]
- Platt JC (1999). Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In *Advances in Large Margin Classifiers* (pp. 61–74). MIT Press.
- Rose NS, LaRocque JJ, Riggall AC, Gosseries O, Starrett MJ, Meyering EE, & Postle BR (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, 354(6316), 1136–1139. 10.1126/science.aah7011 [PubMed: 27934762]
- Samaha J, Sprague TC, & Postle BR (2016). Decoding and Reconstructing the Focus of Spatial Attention from the Topography of Alpha-band Oscillations. *Journal of Cognitive Neuroscience*, 28(8), 1090–1097. 10.1162/jocn_a_00955 [PubMed: 27003790]
- Serences JT, Ester EF, Vogel EK, & Awh E (2009). Stimulus-Specific Delay Activity in Human Primary Visual Cortex. *Psychological Science* (Wiley-Blackwell), 20(2), 207–214. 10.1111/j.1467-9280.2009.02276.x
- Stokes MG (2015). ‘Activity-silent’ working memory in prefrontal cortex: a dynamic coding framework. *Trends in Cognitive Sciences*, 19(7), 394–405. 10.1016/j.tics.2015.05.004 [PubMed: 26051384]
- Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, & Duncan J (2013). Dynamic Coding for Cognitive Control in Prefrontal Cortex. *Neuron*, 78(2), 364–375. 10.1016/j.neuron.2013.01.039 [PubMed: 23562541]
- Treisman A, & Zhang W (2006). Location and binding in visual working memory. *Memory and Cognition*, 34, 1704–1719. [PubMed: 17489296]
- Vogel EK, & Machizawa MG (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428, 748–751. [PubMed: 15085132]
- Vogel EK, Woodman GF, & Luck SJ (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 92–114. [PubMed: 11248943]

- Woodman GF, & Vogel EK (2008). Selective storage and maintenance of an object's features in visual working memory. *Psychonomic Bulletin & Review*, 15(1), 223–229. 10.3758/PBR.15.1.223 [PubMed: 18605507]
- Ye C, Hu Z, Ristaniemi T, Gendron M, & Liu Q (2016). Retro-dimension-cue benefit in visual working memory. *Scientific Reports*, 6, 35573 10.1038/srep35573 [PubMed: 27774983]
- Yu Q, & Shim WM (2017). Occipital, parietal, and frontal cortices selectively maintain task-relevant features of multi-feature objects in visual working memory. *NeuroImage*, 157, 97–107. 10.1016/j.neuroimage.2017.05.055 [PubMed: 28559190]

Figure 1.

Behavioral task trial sequence (stimuli not depicted to scale). Participants were presented with a two-feature object in the center of the screen, and, in different blocks, were asked to remember its orientation, its color or both color and orientation. Each trial started with the presentation of a fixation cross, followed by a 200 ms presentation of the sample display. The sample was followed by 1800 ms long blank delay interval during which participants were asked to remember the sample stimulus. At test, participants were asked report whether a change in the task-relevant feature had occurred by pressing a key.

Figure 2.

Classifier accuracy for task-relevant feature dimension decoding. (A) Time-series decoding of task-relevant feature dimension. Classifier accuracy was consistently above chance throughout the whole trial interval. Thick colored bar at the top of the plot depicts significant time points (corrected for multiple comparisons using permutation testing). Gray bar at the bottom of the plot corresponds to stimulus presentation. Dashed line represents empirically derived chance. Shaded areas correspond to ± 1 SEM. (B) Bar-plot of classifier accuracy across three trial intervals: pre-stimulus (−200–0 ms), encoding (0–300 ms) and delay (300–

2000 ms). Classifier accuracy was significantly above chance during the entire trial interval including the pre-stimulus interval; however, there was a significant increase in accuracy following stimulus presentation and during the delay interval. Error bars correspond to ± 1 SEM. (C) Time-frequency plot of classifier accuracy. Above-chance classifier accuracy was supported by patterns of oscillatory activity across a wide range of frequencies (all time-frequency points significant $p < .001$; corrected for multiple comparisons using permutation testing). Activity in the lower frequency range (4–15Hz) mainly supported decoding accuracy time-locked to the presentation of the stimulus, whereas higher frequencies (>15 Hz) supported decoding at the level of experimental blocks.

Figure 3.

Classifier accuracy for feature identity decoding from evoked power (feature bin 1 vs. feature bin 2). Classification of orientation and color is depicted in the left (ABC) and right (DEF) panels, respectively. (A, D) Classifier accuracy for feature identity classification across time. Feature identity was decodable during the initial ~700 ms of the trial irrespective of feature relevance. Representation of the task-relevant feature persisted the longest. Thick colored bars at the top of the plots depict significant time points (corrected for multiple comparisons using permutation testing). Gray bars at the bottom of the plots correspond to stimulus presentation. Dashed lines represent empirically derived chance. Shaded areas correspond to ± 1 SEM. (B, E) Bar-plot of classifier accuracy across three trial intervals: pre-stimulus (–200–0 ms), encoding (0–300 ms) and delay (300–2000 ms). Classifier accuracy was higher during encoding compared to both pre-stimulus and delay intervals. There was no significant interaction between time interval and feature relevance.

Error bars correspond to ± 1 SEM. (C, F) Time-frequency decoding of feature identity. Above chance classification of feature identity was supported by a wide range of frequencies (orientation 4–25 Hz, color 4–30 Hz). Time-frequency points with $p > .001$ are masked with semi-transparent white color (corrected for multiple comparisons using permutation testing).

Figure 4.

Classifier accuracy for feature identity decoding from total power (feature bin 1 vs. feature bin 2). Classification of orientation and color is depicted in the left (ABC) and right (DEF) panels respectively. (A, D) Classifier accuracy for feature identity classification across time. Classification of orientation was reliably above chance only when orientation corresponded to the task-relevant feature (orientation relevant and both relevant conditions). Decoding of orientation identity in orientation-irrelevant condition was only successful during the initial ~500 ms. Classification of color was only reliable during the initial ~500 ms irrespective of color relevance. Thick colored bars at the top of the plots depict significant time points (corrected for multiple comparisons using permutation testing). Gray bars at the bottom of the plots correspond to stimulus presentation. Dashed lines represent empirically derived chance. Shaded areas correspond to ± 1 SEM. (B, E) Bar-plot of classifier accuracy across three trial intervals: pre-stimulus (-200–0 ms), encoding (0–300 ms) and delay (300–2000

ms). There was a significant interaction between interval and feature relevance for orientation. Classifier accuracy was higher during encoding and delay intervals when orientation was task-relevant compared to task-irrelevant. There was no significant interaction between interval and feature relevance for color. Error bars correspond to ± 1 SEM. (C, F) Time-frequency decoding of feature identity. Above chance classification of orientation identity was supported by patterns of oscillatory power in the alpha-band range (8–14 Hz) across the delay interval. Time-frequency points with $p > .001$ are masked with semi-transparent white color (corrected for multiple comparisons using permutation testing).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 5.

Comparison of classifier evidence between different conditions. Classifier evidence for orientation and color is depicted in the left (AB) and right (CD) panels respectively. Gray bars at the bottom of the plots correspond to stimulus presentation. Thick colored bars at the top of the plots depict time points with significant differences between conditions (corrected for multiple comparisons using Benjamini-Hochberg correction). Dashed lines represent $P = 0.5$. Shaded areas correspond to ± 1 SEM. (A, B) Classifier evidence calculated from evoked power. There was a significant difference in classifier evidence between orientation relevant and orientation irrelevant conditions during the initial 100–600 ms (black thick line at the top of the plot). There was also a significant difference in classifier evidence between conditions where orientation was the only task relevant feature and where orientation was remembered together with color (gray bar at the top of the plot). No significant differences between conditions were observed for color. (C, D) Classifier evidence calculated from total power. Similar to evoked power classifier evidence, there was a significant difference between orientation relevant and orientation irrelevant conditions during the initial encoding of the stimulus. No significant differences between conditions were observed for color.