Maintenance of Bound or Independent Features in Visual Working Memory is Task-dependent

Ruoyi Cao[1*], Yoni Pertzov[2], Zaifeng Gao[3], Mowei Shen[3] , Leon Y. Deouell [1,2*]

[1]Edmond and Lily Safra Center for brain sciences, and [2]Department of Psychology, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel

[3]Department of Psychology and Behavioral Sciences, Zhejiang University, Hang Zhou 310007, P.R China.

*Corresponding

Edmond and Lily Safra Center for Brain Sciences
The Suzanne and Charles Goodman Brain Sciences Building
Edmond J. Safra Campus, The Hebrew University
Jerusalem, 9190401

Tel:+972-2-5881739

Leon Y. Deouell

E-mail: leon.deouell@mail.huji.ac.il

Ruoyi Cao

E-mail: ruoyi.cao@mail.huji.ac.il

Abstract

Over the last decade, seemingly conflicting results were obtained regarding the question of whether features of an object are stored separately, or bound together, in working memory (WM). Many of these studies are based on an implicit assumption about a default, or fixed, mode of working memory storage. However, according to recent findings about the functional property of WM, we proposed that anticipated memory probes used in a given experiment might actually determine the format in which information is maintained in WM. In order to test this flexible maintenance hypothesis, we recorded EEG while subjects performed a delayed-match-to-sample task with and without the requirement of maintaining bound features. In two experiments, we found that EEG signals recorded in central-parietal channels differ significantly between the two conditions, providing reliable evidence for such flexible maintenance.

Introduction

The question of how feature binding is maintained in working memory has received a lot of attention in the psychological and neuroscientific literature (Treisman & Gelade, 1980; Treisman, 1998;Schneegans & Bays, 2019a; Treccani, 2018). These studies focused on investigating two questions: 1) whether objects are stored in the visual working memory as bound objects or separated features, and 2) whether the maintenance of binding in working memory requires the involvement of additional processing resource, for instance, attention. These two questions are interrelated yet independent.

Assessing the limitations of working memory performance is a typical, but indirect, way to investigate the first question concerning the format of objects maintained in working memory, either as bound objects or as separate features. If features are separately stored in working memory, then performance should be limited by the number of features rather than the number of objects to be maintained. Otherwise, if bound objects rather than separate features are stored, then working memory performance should be limited by the number of objects regardless of the number of features. Following such reasoning, Luck and Vogel (1997) proposed a 'strong object' view since they observed that increasing numbers of features within objects did not impair performance in change-detection tasks. Moreover, they found that remembering additional features did not decrease performance compared to remembering less features with the same number of objects. Invariance to the number of features was observed even when one object consisted of multiple values from the same feature dimension. Such a strong object hypothesis was supported by a study showing no difference in delay EEG activity between memorization of multiple and single color objects (Luria & Vogel, 2011).

However, other studies came to different conclusions. For example, Wheeler and Treisman (Wheeler & Treisman, 2002) observed that while features from different dimensions could be stored in parallel without a cost for working memory performance, adding features from the same feature dimension limited memory performance. They proposed the multiple-resources view, stating that working memory maintains features from different feature dimensions in parallel, while features from the same feature dimension compete for storage space. The multiple-resources view was later supported by other studies using either a similar (Delvenne & Bruyer, 2004) and different paradigms (Wang, 2017, Xu, 2002). For example, Wang et al (2017) found that for objects with two dimensions, the memory performance decreased as more feature values had to be remembered, but the other (fixed) dimension of that object was not affacted.

Yet another variant of a proposed architecture of working memory suggests that working memory performance is limited by both the number of objects and the number of features. For example, Olson and Jiang (2002) found that when the number of objects was held constant, performance was better in the single-feature condition than in the multiple feature condition, suggesting that it was more difficult to store two features than one feature of an object. When the number of features was held constant, performance was better when features conjoined to form objects than when

they were presented as isolated features. The importance of both the number of features and the number of items inspired the ‘weak-objects view’ of working memory limitation (Alvarez & Cavanagh, 2004; Hardman & Cowan, 2015).

The second question, whether binding features together requires additional resources, was widely studied through dual-task paradigms in which a secondary-attention-demanding task was performed during the retention interval of a task. If maintaining bound objects requires processing resources beyond those required for maintaining unbound features, then tests requiring retention of bound features would suffer from the secondary task to a greater extent than those requiring only retention of features. Using this approach, Allen, Baddeley and Hitch (2006) found that a concurrent backward counting task impaired overall change-detection performance, but did not impair memory of bound objects more than memory of separate features. This pattern of results was replicated when memory of bound objects was made more demanding by separating the shape and color features spatially, temporally, or across visual and auditory modalities (Baddeley, Allen, & Hitch, 2011), and when the secondary task was auditory (Morey & Bieler, 2013; Vergauwe, Langerock, & Barrouillet, 2014). That is, the impairment caused by the secondary task was not selective to bound objects.

However, some secondary tasks, involving object-based attention, like multi-object tracking (Fougnie & Marois, 2009), mental rotation, and random dot kinematograms (He et al., 2020; Shen, Huang, & Gao, 2015), showed a greater interference with the maintenance of bound objects. A specific susceptibility of feature bindings to interference from subsequent distractor stimuli has also been observed (Makovski & Pertzov, 2015; Ueno, Allen, Baddeley, Hitch, & Saito, 2011). A fMRI study showed that retention of binding in addition to features involved more cortical regions compared with retention of features (Parra, Della Sala, Logie, & Morcom, 2014). Behavioral dissociation between maintaining features and bound objects was observed in different groups. For example, memory for bound objects could be impaired, without impairment in memory for individual features or locations, in patients with a variant of encephalitis (Pertzov et al., 2013) and Alzheimer’s disease (Liang et al., 2016; Parra et al., 2009). This suggests that maintenance of binding involves extra processing resources in addition to maintenance of features.

Both debates about whether features are stored separately or bound, and whether keeping bound objects in WM requires extra attention or not, are based on an implicit assumption that there is a default or fixed mode of working memory storage. Under this assumption, this mode of representation could be impaired with resource shortage and could be estimated with performance in response to different probe questions. However, such a universal default mode cannot be taken for granted. Indeed, remarkable flexibility in prioritizing information in WM according to the task goal has been reported. For example, when non-human primates viewed the same to-be-remembered stimuli but were trained to expect different kinds of memory probes, delay activity in the prefrontal cortex showed different patterns (Rao, Rainer, & Miller, 1997). In an fMRI study in humans, participants were required to remember a face and a scene (Nobre, 2007). During the delay period, a cue was presented to inform which is to be tested later. Increased activity was observed in areas involved in

face (fusiform gyrus) or scene (parahippocampal gyrus) processing, according to the cue. Indeed, working memory is not simply a passive representational state of visual input during a delay period, but is better conceived as a functional state bridging previous contexts and sensations to anticipated actions and outcomes (Myers, Stokes, & Nobre, 2017). Consequently, the anticipated memory probes (questions) used in a given experiment might actually determine the format in which objects will be maintained in working memory and the involved cognitive resources.

Imaging studies have also supported both maintenance of separate features and of bound objects. For instance, both primate (Baizer, Ungerleider, & Desimone, 1991; Mishkin, Ungerleider, & Macko, 1983) and human studies (Courtney, Ungerleider, Keil, & Haxby, 1996; Smith et al., 1995) found that working memory for spatial location and item identities activated different regions of the brain. Other studies however found no evidence for distinct representations of spatial and non-spatial features (D'Esposito et al., 1998; Kravitz, Kriegeskorte, & Baker, 2010; Kravitz, Saleem, Baker, & Mishkin, 2011). These conflicting findings might be due to different methodologies used, but they may also be due to the flexibility of storing visual input both as bound objects and as separated features.

In order to test this flexible maintenance hypothesis, claiming that the brain stores separate features or bound objects according to the task goal, we recorded EEG while subjects performed a delayed-match-to-sample task with and without the requirement of maintaining binding between features. If there is a universal default mode for the system to store visual stimuli regardless of task requirement, we expected no systematic differences between these two conditions in the Event-Related Potential (ERP) during delay period, as the visual input to be maintained was identical. If information maintained in the visual working memory is adaptive to task goals, we expected to find differences in ERP during the delay period between these two conditions. The direction of such differences, if any, could inform us of the amount of processing resources consumed by each representation format. We report two experiments, where Experiment 1 is exploratory, and Experiment 2 is a confirmatory replication.

Experiment 1
Methods

Participants

Fifteen healthy volunteers from the Hebrew University of Jerusalem participated in the study. They were paid (40NIS/hrs.) or given course credits for participation. All subjects had reportedly normal or corrected-to-normal sight and no psychiatric or neurological history. One subject did not finish the experiment and three were excluded from analysis due to noisy recordings. The remaining eleven subjects consisted of 6 males and 5 females 19–31 years old. The experiment was approved by the ethics committee of the Hebrew University of Jerusalem, and informed consents were obtained after the experimental procedures were explained to subjects.

Stimuli and apparatus

Subjects sat in a dimly lit room. The stimuli were presented with Psychotoolbox-3 (http://psychtoolbox.org/) implemented in Matlab 2018 on a ViewSonic G75f CRT (1024×768) monitor with a 100-Hz refresh rate. They appeared on a grey background at the center of the computer screen located 100 cm away from the subjects' eyes.

Subjects performed a delayed match-to-sample test with 3 blocks of trials from 3 conditions (Figure 1): two-item-feature (F2), two-item-binding (B2), and four-item-binding (B4). In the B2 and F2 conditions the memory array consisted of 2 colored items randomly selected out of six items of identical irregular shapes in clearly different colors , which are red (250 ,0 , 0),green (0, 250, 0), yellow (250 ,250, 0),blue (0, 0, 250),violet(250,250,0)and white(0,0,0) (Figure 1). Each item subtended a visual angle of 2.1°×2.1. The color of each item was randomly selected out of six highly distinguished colors, without repetition (i.e. each item had a unique color within the array). The locations of the two selected items were randomly selected out of eight potential locations evenly distributed on an invisible circle with a diameter of 7.3° centered on the fixation cross. In the B4 condition, the memory array consisted of four items of different colors on four locations that were randomly selected from the six colors and eight potential locations respectively, without repetition.

Following a variable delay period of 1, 2 or 3 seconds, a single probe was shown on each trial. In the B2 and B4 condition, a probe could be one of three types: *Matched*, *New-Feature*, or *Mis-Conjunction*. A *Matched* probe had the same color and location as one of the items in the memory array. A *New-Feature* probe had either an old color but was located at a novel location, or was at an old location but had a novel color. *Mis-conjunction probes* had an old color and an old location, both present in the memory array, but the conjunction between them was novel. In the F2 condition, there were only two types of probes, namely Matched probes and New-Feature probes. A *Matched probe* in this condition was either presented at the center of the screen in an old color contained in the memory array, or a black item shown at an old location. Finally, a *New-feature probe* in this condition was a probe either presented at the center of the screen with a color that did not appear in the previous memory array or an item in black presented at a new location.

Experimental procedure

Each trial started with 700 msec a fixation cross presented at the center of the screen. Next, the memory array appeared for 100 msec. This memory array was followed by delay period of either 1,2 or 3 seconds blank screen, in equal proportions. Trials with the 3 delay durations were randomly mixed within a block. Following the delay period, a probe item was presented and subjects had to press a key to indicate if the probe item was an "old" or a "novel" item. The probe disappeared when the response was made (with a limit of 2 seconds), followed by the initial fixation of the next trial.

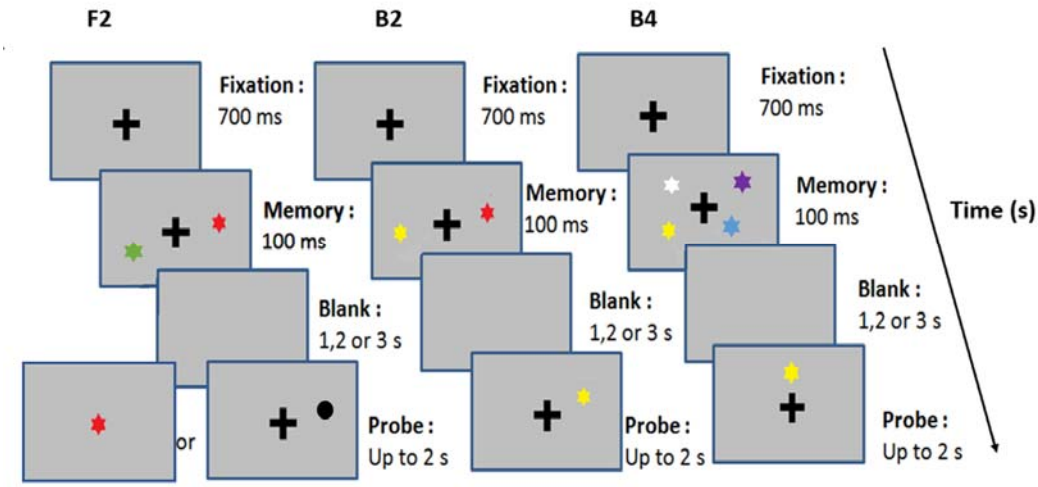

Figure 1. Illustration of an example trial in the three conditions: Left – Two-Item feature condition (F2) with Matched probe , middle – Two-Item binding condition (B2) with Mis-conjunction probe , right – Four-Item binding condition (B4) with New-feature probe.

There were 3 conditions in the experiment, each consisting of 288 trials. In the B2 and B4 conditions, the delay period was followed by match probes in 144 trials, new feature probes in 48 trials, and *Mis-conjunction* probes in 96 trials, and a probe was regarded as an old item only when the probe had both the color and the location of one of the items shown in the memory array. In the F2 condition, the delay period was followed by *match-feature* probes in 144 trials and the *new-feature* probes in the remaining 144 trials, and a probe was regarded old when either the color or the location of the probe had been shown in the memory array. Each condition began with 64 practice trials and each condition was divided into 2 consecutive blocks, resulting in 6 blocks in total. Each block took about 15 mins, and subjects were instructed to take a break between blocks.

As the memory arrays in B2 and F2 both included 2 items with 2 relevant features each, the contrast between B2 and F2 was intended to reveal the effect of retaining feature conjunctions in addition to individual feature values. The contrast between B4 and B2 was conducted to reveal a load effect (maintaining 4 vs. 2 items). The order of these three conditions was counterbalanced across participants.

All analyses and figures for the behavior results were made using JASP 0.8.5.1 (https://jasp-stats.org/). Response accuracy of 11 subjects was entered into a 3 (Delay duration: 1 vs 2 vs 3 sec delay) × 3 (Condition: F2 vs B2 vs B4) repeated-measures two-way ANOVA. Degrees of freedom were adjusted for violations of the assumption of sphericity with the Greenhouse–Geisser correction when necessary.

EEG recording and data analysis

EEG data were recorded using an Active 2 system (BioSemi, the Netherlands) from 64 active electrodes spread out across the scalp according to the extended 10–20 system with the addition of two mastoid electrodes and a nose electrode (https://www.biosemi.com/pics/cap_64_layout_medium.jpg). Horizontal electro-oculogram (EOG) was recorded from electrodes placed at the outer canthi of both

eyes. Vertical EOG was recorded from electrodes placed above and below the left eye. The EEG was continuously sampled at 1024 Hz with an anti-aliasing low pass filter with a cutoff of 1/5 the sampling rate, and stored for off-line analysis. The data was referenced online to the Common Mode Sensor (CMS) which was placed in the space between POz, PO3, Pz, and P1 .

Data preprocessing and analysis were done with the FieldTrip toolbox (version 20191213 http://www.fieldtriptoolbox.org/) implemented in Matlab 2018 (Mathworks, Natick, MA, USA). Preprocessing was applied to continuous data. During preprocessing, EEG and EOG signals were firstly filtered with Butterworth zero-phase (forward and reverse filter) Bandpass filter of 0.1–180 Hz and then referenced to the nose channel. Extremely noisy or silent channels, which contributed more than 20% of all artifacts (Criteria: more than 100μV absolute difference between samples within segments of 100 ms; absolute amplitude > 100μV) were deleted. No more than 2 channels in a single subject were deleted, see details below. Next, data were re-referenced to an average of all remaining EEG electrodes. Ocular and muscular artifacts were removed from the EEG signal using the ICA method by manual selection of artifact components based on correlation with the EOG channels, power spectrum typical to muscle activity, and typical component scalp topographies. After ocular and muscle artifacts were removed, automatic artifact rejection was applied (http://www.fieldtriptoolbox.org/tutorial/automatic_artifact_rejection/). Time points larger than 12 standard deviation from the mean of the corresponding channel were marked, together with 200 msec before and after, ensuring that the (subthreshold) beginning and end of an artifactual event will be accounted for. A visual inspection of the data followed in order to detect rare artifacts which were missed by the automatic procedure. Finally, previously deleted channels were recreated by mean interpolation of the neighboring electrodes (FC5, F6 were interpolated in subject 01, AFz in subjects 02, 04, and 06, AF8 and Iz in subjects 05, 07, and 09, and PO3 and CP1 in subject 08). Data were then down-sampled to 512 Hz, filtered with a Butterworth zero-phase lowpass filter at a cutoff at 20 Hz and parsed into 1800 msec segments starting 500 msec before the memory array onset, and averaged within each subject and condition. The average of 100 msec before the onset of stimuli over each trial was defined as baseline and subtracted from all data points of each segment.

There were two planned contrasts in our experiments: (a) the difference between F2 and B2, potentially revealing the effect of task (retaining features or bound objects) on the ERPs, given the same visual input, and (b) the difference between B2 and B4, revealing the working memory load effect on the ERP, given the same task. To compare ERP amplitudes between conditions across all electrodes and time samples, cluster-based permutation tests (Maris & Oostenveld, 2007) were performed in Experiment 1, the exploratory phase. This approach allowed for a sensitive comparison between conditions at the level of spatiotemporal clusters without a predefined region of interest (ROI) and provided relevant correction for multiple comparisons with Monte Carlo based cluster-correction. The cluster-based permutation test included the following procedure: a) A paired sample t-test was applied to each time point and channel, resulting in one t-value for each data point. b) A threshold of $p < .05$ was applied to each time point on each channel, and a cluster

was defined as the collection of above threshold data points, adjacent to each other either spatially or temporally. c) t-values were summed up within each cluster resulting in sum-t values. d) Repeating steps a-c for 1000 times on data while switching the condition labels within a randomly selected set of subjects for each iteration. Within each iteration, the largest positive and the smallest negative sum-t entered into the “positive” and the “negative” null distribution, respectively. e) The sum-t value of each cluster calculated in step c was compared with the two null distributions. Clusters with the sum t-value larger than of 97.5% of the largest null distribution were defined as significant positive clusters, while clusters with the sum-t value smaller than 97.5% of the smallest null distribution were defined significant negative clusters.

In addition to directly contrasting ERPs between conditions, multivariate pattern analysis (MVPA) was applied to test whether we can predict the condition based on the topographic patterns of brain activity at each time point. In addition to training and testing classifiers to distinguish B2 vs F2 contrast and separately B4 vs B2 contrast, we used cross-condition decoding to check whether the discriminating pattern revealed by the contrast between B2 vs B4 shared the same underlying pattern as revealed by B2 vs F2 contrast. Finding significant cross-contrast decoding would suggest shared processes, possibly reflecting differences in task load in both contrasts.

MVPA was performed in MATLAB using the MPVA Light toolbox (https://github.com/treder/MVPA-Light). The classification was performed for each participant in a time-resolved manner. For each participant, trials in each condition were first randomly sub-sampled to the minimum number within all three conditions to attain a balanced set for training and testing. These data were then split into five equal parts. Then, a linear discriminant (LDA) classifier was trained for each time point from 200 msec before the stimuli onset to 1100 msec after the stimuli onset, using the voltage at 64 electrodes as features, using 5-fold cross-validation. For each fold, the classifier was trained repeatedly on 80% of the data and tested on the remaining 20% of trials. To increase the signal-to-noise ratio, training and testing were done on averages over 4 trials selected randomly with replacement within training and testing sets separately. The final results were averaged over the five folds. To attain decoding performance invariant to the decision boundary, Area Under Curve (AUC) was calculated for each subject. To reveal the generalization of these resulting classifiers over time, a temporal generalization matrix was computed for each subject, in which classifiers were trained on one-time point, and tested on all other time points. This provided detailed information about the sequence of processing stages engaged in the task. The above decoding procedures were applied on decoding two within contrast pairs (B2 vs B4 and F2 vs B2) to show the load effect and the binding effect respectively, and two across contrast pairs (Training on B2 vs B4, Testing on F2 vs B2; Training on F2 vs B2, Testing on F2 vs B2) to show the independency (or dependency) of these two effects.

For each decoding analysis, a cluster-based permutation test was conducted to compare AUC with 0.5 with correction for multiple comparisons at the group level ((Maris & Oostenveld, 2007)). That is, the AUC of all subjects for each time point was first compared with 0.5 chance level with one-sample-test and then threshold at

0.01 significant level. Positive significant t-values of data points adjacent in time that survived the first level threshold defined clusters, and the t-values within each cluster was summed up to create t_{sum} values for each cluster (note that in the temporal generalization matrix, adjacent in time is defined as either adjacent in testing time or in training time). The t_{sum} values was compared with the distribution of the largest t_{sum} values in 1000 iterations generated by the same procedure, albeit with the chance level (AUC=0.5) and the true decoding performance switched in label for a random subsample of the subjects.

Results

Behavioral results

A 3 (Delay duration : 1 sec vs 2 sec vs 3 sec delay) × 3 (Condition: 2F vs 2B vs 4B) repeated-measures two-way ANOVA of response accuracy revealed a main effect of Condition, $F(2,20) = 13.70$, $p < .001$, $\eta 2 = .58$ (Figure 2). Follow-up pairwise comparison (with Bonferroni correction) showed a significantly lower response accuracy in condition B4 than both condition B2 , *Mdiff* = - .095, *SE* = .03, $t = -3.30$, $p = .011$, *Cohen'd* =-.996, and condition F2, *Mdiff* = - .15, *SE* = .03, $t = -5.18$, $p < .001$, *Cohen'd* = -1.56, whereas no significant differences were found between B2 and F2 condition on the response accuracy, *Mdiff* = - .054, *SE* = .03, $t = -.56$, $p = .23$. The main effect of Delay Duration was also significant, $F(2,20) = 4.89$, $p = .002$, $\eta 2 = .328$. Follow-up pairwise comparisons showed a marginally significant decrease of response accuracy from the 1 sec to the 3 sec delay, *Mdiff* = .03, *SE* = .01, $t = 2.49$, $p = .065$, *Cohen'd* = .75, and the 2 sec to the 3 sec delay, *Mdiff* = - .026, $t = 3.96$, *SE* = .007, $p = .006$, *Cohen'd* = .75, but not from 1 to 2 sec delay, *Mdiff* = - .003, *SE* = .01, $t = -.26$, $p = 1$, reflecting some memory decrement following extended delays. The interaction between Delay Duration and the Condition was not significant, $F(4,40) = .69$, $p = .60$, $\eta 2 = .07$. These results indicated that the B4 condition, as expected, was more difficult than the other conditions, and that F2 and B2 conditions did not significantly differ in their overall exertion. Therefore, any results gained from comparing F2 and B2 conditions in EEG could not be explained by differences in task difficulty.

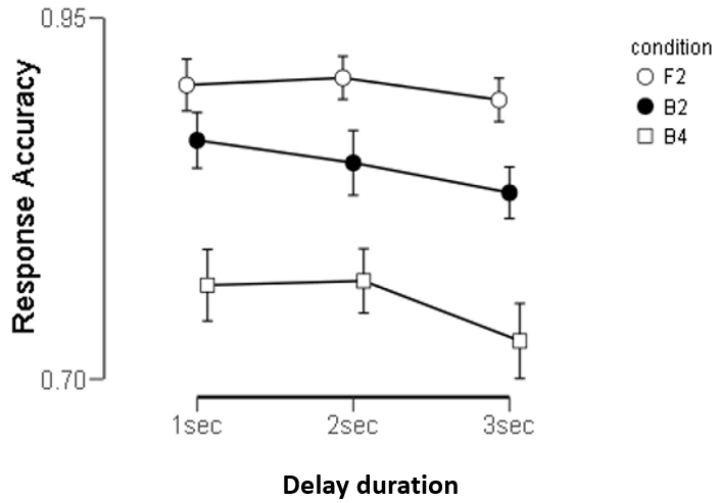

Figure 2: *Percentage of correct responses on each condition following 1, 2, or 3sec delay between the memory and test. Error bar represents standard error across subjects.*

ERP results

Load effect: comparing B4 and B2 condition

We examined the effect of enhanced memory load by comparing the EEG signal amplitude in the B4 and B2 conditions up to the first second after the onset of the memory array (as this period was common to all 3 delays). Figure 3a shows that maintaining four-bound objects (B4 condition) evoked larger ERP amplitude than maintaining two-bound objects (B2 condition) at right-frontal and middle-central electrodes. A spatiotemporal cluster-based permutation test showed that this significant cluster of difference ($p < 0.05$, cluster-corrected) extended from 394 msec to 847 msec after the onset of the memory array[1]. Figure 3b shows the grand average waveforms across subjects for 25 channels included in this significant positive cluster[2].

Binding effect: comparing F2 and B2 condition

We compared ERP amplitude when subjects were required to remember two items with (B2) and without (F2) the need to maintain the conjunction (binding) between color and location. As the memory array was identical in both conditions, a difference between the two conditions could not be due to differences in visual information. This contrast thus indicated whether the visual information had been maintained in a task-specific manner. Cluster-based permutation test showed that F2 elicited stronger positivity than B2 with a significant cluster in centro-parietal electrodes between 220 msec to 589 msec after the onset of the memory array (Figure 3c). Figure 3d shows the amplitude averaged across subjects and 20 channels included in this significant positive cluster.

---

[1]Note that the cluster analysis does not ensure that the effect in any single point in the cluster is independently significant (Sassenhagen & Draschkow, 2019)

[2]Note that a negative cluster on the left-temporal region is occasionally significant (P<.05), but it was not stable across multiple run in permutation.

MPVA results

Multivariate analysis further revealed different responses between tasks. The temporal generalization matrices averaged over subjects are shown in Figure 3e. Cluster-based permutation test detected a significant cluster when training and testing on B4 vs B2 condition (Figure 3e bottom-right panel, $P = 0.006$, cluster-corrected). Figure 3f shows the diagonal of the temporal generalization matrix (Figure 3e), where testing and training were conducted on the same time points. Above chance decoding was found from 6 msec after stimulus onset until end of the maintenance period with the peak accuracy at 152 msec ($P = 0.007$ cluster corrected). Although the significant cluster started early after the stimuli onset, classifiers trained on early time points could not decode the later maintenance period above chance. Meanwhile, the later component starting around 200 msec after the stimuli onset generalized to the end of the maintenance period. This suggests that the significant contrast in B2 vs B4 found in the above ERP analysis reflected a working memory load effect, which is different from the early visual effect.

Despite the identical visual presentation, training and testing on the same time point showed a significant decoding performance (AUC) between B2 and F2 from 111 msec to 630 msec with the peak accuracy at 259 msec ($P < 0.001$ cluster corrected). This above chance decoding reflects a “binding effect”. The temporal generalization matrix indicated a relative stable temporal dynamic starting from around 200 msec after the stimuli onset and was lasting until around 800 msec (Figure 3e Top-left panel ,$P = 0.068$, cluster-corrected).

Next we asked whether we could decode B2 and F2 using a classifier trained on discriminating B2 and B4 and vice versa. Such cross decoding would suggest shared distributions between “load effect” and the “binding effect”. However, despite the fact that we could decode B2 from F2 and B2 from B4 significantly above chance, we were not able to find a significant above chance cluster, neither when training and testing at the same time points (diagonal) or across the generalization matrix. This was the case whether we trained on the B2 vs F2 contrast and tested on the B2vs B4 contrast or vice versa. This is consistent with the observation that the spatial distribution of the B2-B4 contrast was not identical to that of the B2-F2 contrast (compare Figures 3a and 3c).

Before discussing the results further, we note that the task-related differences were found through an exploratory method rather than a hypothesis-driven approach. Therefore, in experiment 2, we recruited a new group of subjects in an attempt to replicate this finding with the same paradigm.

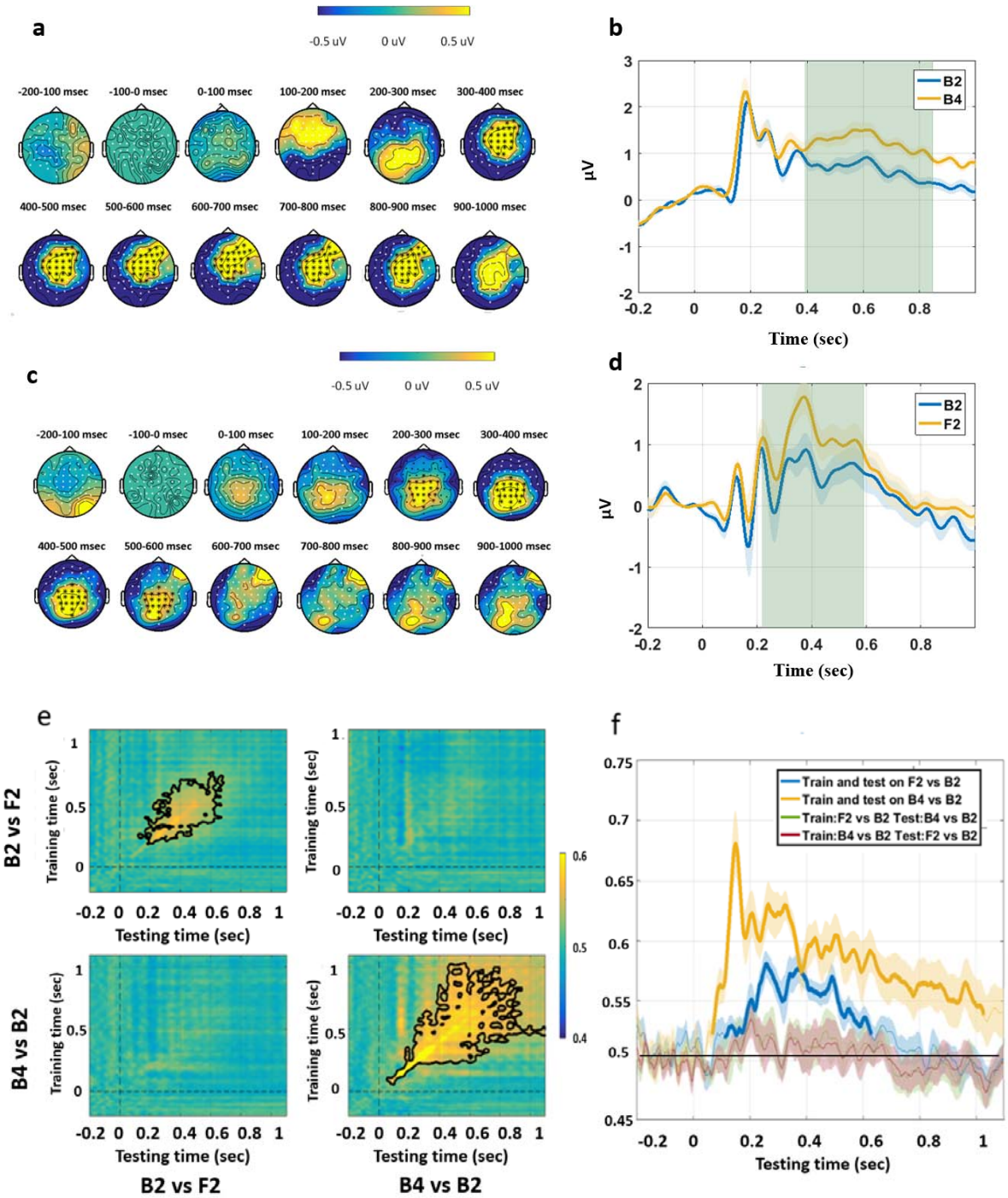

Figure 3 a*: Difference in topographies of response amplitudes between B4 and B2 condition, in 100 msec bins, from 200 msec before until 1sec after the memory array onset. Yellow colors show higher positivity for B4 than B2. Black markers stand for channels included in significant clusters based on the cluster permutation procedure.* ***b:*** *ERPs elicited by the B4 and B2 conditions, averaged over subjects and channels included in the significant cluster. The light green rectangle highlights the temporal duration for the "Rectangle ROI" (see Figure 5a &Figure 5a). Shading around the waveforms represents standard error of the mean across subjects. 0 on the X-axis corresponds to the onset of memory array.* ***c****: Difference topographies between F2 and B2 conditions, in 100 ms bins from 200 ms before memory array onset. Yellow color indicates higher positivity in F2 conditions than in B2 condition while blue colors indicate the opposite. Black markers stand for channels included in significant clusters.* ***d:*** *As in* ***b****, ERPs elicited in*

the F2 and B2 conditions, averaged over subjects and channels in the significant cluster. **e**: *Temporal generalization matrices. Training time is shown on the vertical axis and testing time on the horizontal axis. Significant clusters are indicated by black contours.* ***f:*** *Decoding performance (AUC) in Experiment 1, for classifiers trained and tested on the same time points (diagonals of the matrices in panel e). Thick lines indicate significant above chance decoding (cluster corrected, p<0.05). The shaded area represents the standard error across subjects.*

Experiment 2
Method

We replicated experiment 1 using the same experimental material and procedure.

Participants
Eighteen healthy volunteers from the Hebrew University of Jerusalem participated for either course credits or payment (40NIS/hrs). All subjects had normal or corrected-to-normal sight and reported no psychiatric or neurological history. Five subjects were excluded from analysis due to noisy recordings (Subjects in which more than two neighboring channels were noisy were excluded). The final sample was composed of thirteen subjects (6 males, 7 females aged range 21-30) [3]. Informed consents were obtained after the experimental procedures were explained to the subjects. The experiment was approved by the ethics committee of the Hebrew University of Jerusalem.

EEG recording and data analysis
Data preprocessing and analysis were the same as in Experiment 1. Briefly, preprocessing started with deleting extremely noisy channels. A Butterworth Bandpass filter of 0.1–180 Hz was applied. Ocular and muscular artifacts were extracted with ICA and removed. Then an automatic and manual artifact rejection was applied. Finally, previously deleted channels were recreated by mean interpolation of the neighboring electrodes (CP3 and P1 in subject 01; AF7, F7, P9 in subject 03; FC5 in subject 04; POz, Fp2 in subject 05; T7,TP8 and P9 in subject 06; F8 in subject 08, CP3, T7 in subject 09 ; T7, CP3 in subject 12) . After preprocessing, data were down-sampled to 512 Hz and were low-pass filtered with a Butterworth bandpass filter at a cutoff of 20 Hz, segmented, and corrected for the baseline.

In Experiment 2, the confirmatory phase, with a new group of subjects, we used pre-defined ROIs based on the significant spatio-temporal clusters found in experiment 1. “Rectangle ROI” for the contrast between two conditions was defined as all channels that were included in the significant cluster, spanning the interval from the earliest time to the latest time in that cluster across all channels. “Cloud ROI “ included all channels within the significant clusters but including, for each channel, only the time points inside the significant cluster of that channel (Figure 5a &5d). One-tailed-paired-sample t-tests were then applied to compare amplitudes averaged over pre-defined ROIs (Rectangle ROI & Cloud ROI) between B2 and B4 conditions and F2

[3]Given the effect size of the B2 and F2 contrast in experiment 1 using the “cloud” ROI (the smallest effect size in all possible contrasts), this number of subjects gave power of 96% with 0.05 alpha level in Experiment 2.

and B2 conditions respectively. The direction of the one-tailed test followed the expected effect as observed in Experiment 1.

Results

Behavior results

A 3 (Delay duration : 1 sec vs 2 sec vs 3 sec delay) × 3 (Condition: 2F vs 2B vs 4B) repeated-measures ANOVA revealed a main effect of Condition on response accuracy, $F(2,24) = 27.38$, $p < .001$, $\eta 2 = .70$. Follow-up pairwise comparisons (all with Bonferroni corrections) showed a significantly lower response accuracy in condition B4 than both condition B2, *Mdiff* = - .15, *SE* = .023, $t = -6.65$, $p < .001$, *Cohen'd* = -1.85, and condition F2,*Mdiff* = - .14, *SE* = .023, $t = -6.13$, *Cohen'd* = -1.70, $p < .001$, while no significant differences were found between B2 and F2 , *Mdiff* = 001, *SE* = .023, $t = 0.52$, *Cohen'd* = 0.14 , $p = 1$. These results are similar to the pattern of behavioral results in Experiment 1 and confirmed that B2 Condition and F2 Condition were not significantly different in terms of task difficulty, while B4 Condition was more tasking than the other two. As in experiment 1, a main effect was found on Delay Duration, $F(2,24) = 7.70$, $p = .003$, $\eta 2 = .39$ (Figure 4). Follow-up pairwise comparison showed a significant decrease of response accuracy following 3 secs delay compared with 1 sec delay , *Mdiff* = - .032, *SE* = .01, $t = -3.91$, $p = .009$, *Cohen'd* = -1.09, while no significant difference in response accuracy was found between 1 sec delay and 2 sec delay, *Mdiff* = -.014, *SE* = .008, $t = -1.74$, $p = .28$, *Cohen'd* = -0.48, nor between 2 sec delay and 3 sec delay, *Mdiff* = - .02, *SE* = .008, $t = -2.17$, $p = .12$, *Cohen'd* = -.62. These results reflect forgetting of visual information over time. The interaction between Delay Duration and Condition was marginally significant, $F(4,48) = 2.56$, $p = .05$, $\eta 2 = .18$. Altogether, behavior results confirmed what was found in experiment 1, suggesting that B4 condition is more demanding than the other two conditions, while any result gained from comparing F2 and B2 conditions in EEG could not be explained by differences in task difficulty.

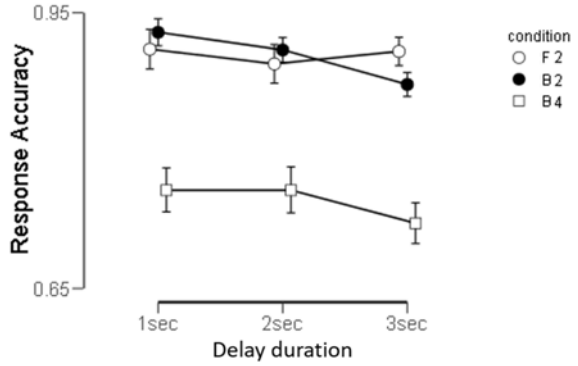

Figure 4: *Percentage of correct responses on each condition following 1, 2, or 3sec delay between the memory and test in Experiment 2. Error bars represent standard error across subjects.*

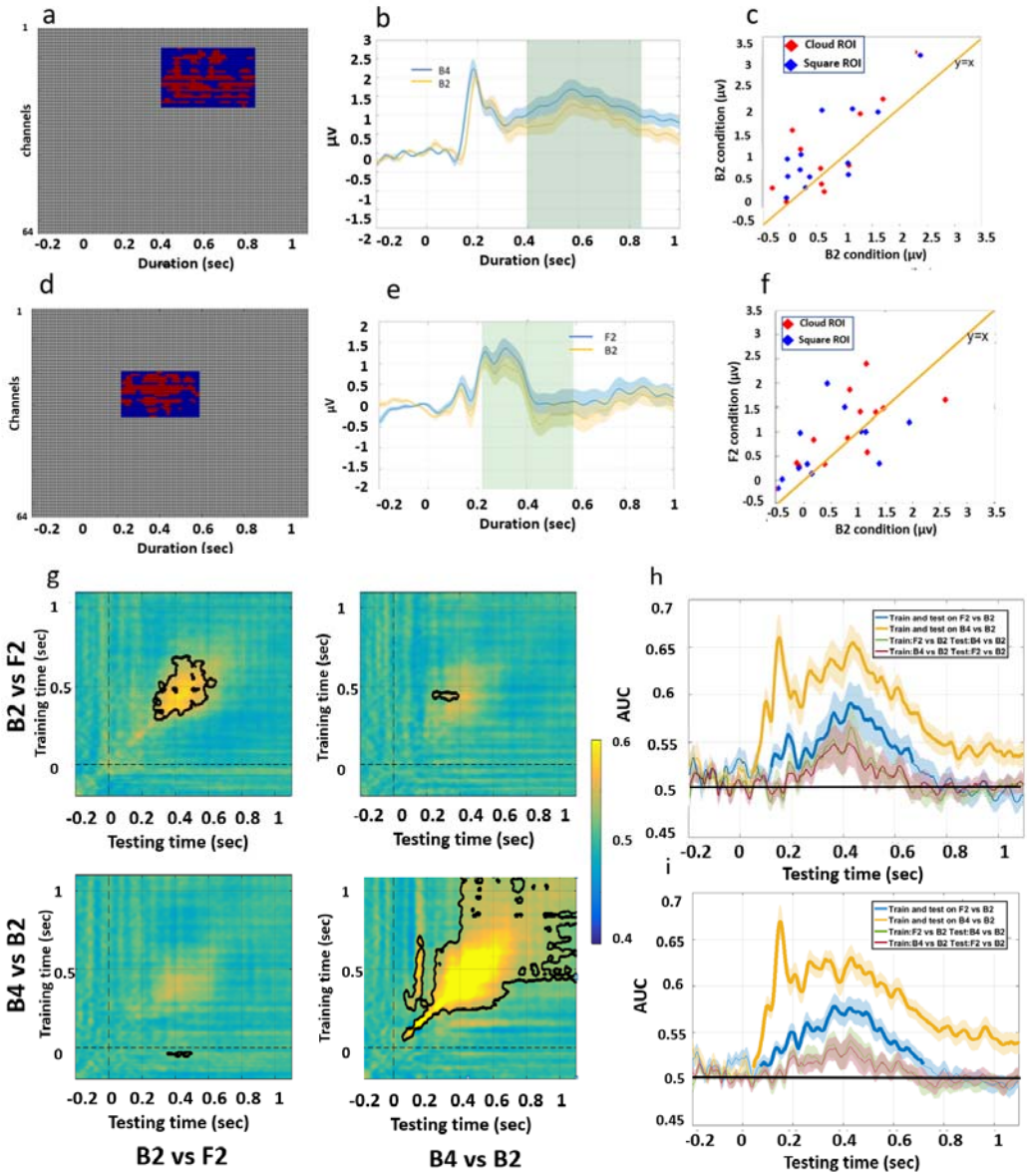

Figure 5 *: **a**:Matrix of channels x time, with red indicting time points and channels included in the significant cluster in the B2 vs B4 contrast in Experiment 1, representing the 'cloud ROI' for this contrast. The rectangle ROI (blue) applies the maximal temporal extent of the significant cluster to all the electrodes included in the cluster. From top to bottom, channels included in the significant clusters are AF4, AF8, F1, Fz, F2, F4, F6, F8, FC3, FC1, FCz, FC2, FC4, FC6, C3, C1, Cz, C2, C4, CP1, CPz, CP2, CP4, Pz, P2* ***b:*** *ERPs elicited by the B4 and B2 conditions, averaged over subjects and channels included in the a priori cluster defined in Experiment 1. The light green rectangle highlights the temporal duration for the "Rectangle ROI". Shading around the waveforms represents standard error of the mean across subjects. 0 on the X-axis corresponds to the onset of memory array.* ***c****: Blue dots and red dots represent the mean amplitudes elicited by B4 and B2 condition averaged over the 'rectangle ROI' and 'cloud ROI' for each subject respectively.* ***d****: As in* ***a****, for the F2 and B2 contrast. From top to bottom, channels included in the significant clusters are Cz, C3, C1, FCz, C2, C4, CP3, CP1, CPz, CP2, CP4, P3, P1, Pz, P2, P4, PO3, POz, PO4, Oz* ***e:*** *As in* ***b****, ERPs elicited in the F2 and B2 conditions, averaged over subjects and channels in the ROI defined in Experiment 1.* ***f****: Blue*

dots and red dots represent the mean amplitudes elicited by B2 and F2 condition averaged over the 'rectangle ROI' and 'cloud ROI' for each subject respectively. ***g****: Temporal generalization matrices from experiment 2. Training time is shown on the vertical axis and testing time on the horizontal axis. Significant clusters are indicated by black contours.* ***h:*** *Decoding performance (AUC) in Experiment 2 when classifiers were trained and tested on the same time points (diagonal of the generalization matrices in g). Thick lines indicate significantly above chance decoding (cluster corrected, p<0.05). The shaded area represents the standard error across subjects. The AUC =0.5 is the chance level.* ***i****:The same as the h but with data in both experiments combined.*

ERP results

Load effect: comparing B4 and B2 condition

Experiment 2 served as a confirmation stage following the exploratory stage of Experiment 1. Thus, the ROIs obtained from Experiment 1 were used to test for Condition Effect in Experiment 2, performed on a new group of subjects. Figure 5b shows the average ERP over subjects and channels included in the ROI. We averaged the amplitudes of the ERPs across the channels for the "Rectangle" ROI (Figure 5c dots in blue) including the channels consisting the significant cluster of Experiment 1, and all the time points within the cluster which were significant in any of the channels. The normality of the data was confirmed with Shapiro-Wilk test. A one-tailed paired sample t-test confirmed that, as in Experiment 1, the B4 condition *(M* = 1.13*, SE* = 0.24*)* evoked significantly higher positivity than the B2 condition *(M = 0.68, SE = 0.20), t(12)* = 3.14*, p* = .004*, Cohen'd* = 0.87. This direction of difference was found in eleven out of thirteen subjects (Figure 5c). A similar result (Figure 5c dots in red) was found using the "Cloud ROI", which was the exact cluster found significant in Experiment 1, by channels and time-points*, t(12) = 3.06, p =.005, Cohen's d = 0.85.* Taken together, we successfully replicated the results showing that maintaining additional bound items in memory lead to larger positive response in frontal-right and middle-central channels within the time window ranging from ~390 msec to ~850 msec after the offset of the memory array.

Binding effect: comparing F2 and B2 condition

The amplitude for F2 and B2 condition were averaged respectively within the spatio-temporal Rectangle ROI (Figure 5d). A One-tailed paired-sample t-test showed that the F2 condition evoked larger positivity (*M = 0.73, SE = 0.18)* than B2 condition (*M* = 0.40, *SE* = 0.22) within the Rectangle ROI. However this difference was not significant, *t(12) = 1.54, p =.08, Cohen's d = 0.43.* Seven out of thirteen subjects exhibited larger averaged positivity in F2 than in B2 condition (Figure 5f, blue dots). A similar result was found when the Cloud ROI was applied (Figure 5f, red dots). A one-tailed paired sample t-test confirmed 2F condition *(M = 1.056, SE = 0.691)* evoked larger positivity than 2B condition, *t*(12) = 1.466, *p* =.086, *Cohen's d = 0.41*.

MPVA results

MPVA further revealed a "load effect" comparing the B2 and B4 contrast. Testing and training on the same time points achieved significant decoding accuracy from 48 msec after stimulus onset and lasted until the end of the maintenance period with the peak accuracy at 154 msec (*P* < 0.001 cluster corrected, Figure 3h). Cluster-based permutation test detected a significant cluster on the temporal generalization matrix (Figure 5g bottom-right panel, *P* = 0.005, cluster-corrected). As found in Experiment

1, a distinguishing pattern between B2 and B4 appeared early on, but only the later component starting around 200 msec generalized to the end of the maintenance period.

A “binding effect” was also found by MPVA. Training and testing on the same time point found a significant above-chance decoding performance (AUC) between B2 and F2 from 123 msec to 671 msec with the peak accuracy at 430 msec ($P < 0.001$ cluster-corrected). The temporal generalization matrix shows a stable significant cluster starting from around 200 msec after the stimuli onset and lasting until around 800 msec (Figure 5g Top-left panel ,$P = 0.069$, cluster-corrected).

Next we asked whether “binding effect” and “load effect” are based on the same mechanism by attempting to discriminate B2 and F2 using a classifier trained on discriminating B2 and B4, and vice versa. Similar to the cross-contrast decoding result in Experiment 1, we were unable to decode across contrasts on either direction on Experiment 2, with the exception of a very short lived off-diagonal cluster shown in figure 5g (black contour in top right panel). No cross-contrast decoding was found when we pooled the subject across experiments as well (Figure 5i).

To conclude, both the “load effect” and the “binding effect” was again showed by successful within-contrast decoding. Around chance cross-contrast decoding performances suggested that these two effects were distinct.

Post-hoc analysis

ERP

The ERP over channels within the pre-defined ROI during the maintenance period reveals a clear trend for F2 condition to evoke larger positive ERP than the B2 possible reason that the difference between F2 and B2 condition, with a somewhat longer delay than in the first experiment (see Figure 6a).A in Experiment 1 did not fully replicate is that the effect differs in the temporal domain across individuals. That is, a large difference between F2 and B2 conditions might appear soon after the offset of the memory array for some subjects, whereas for other subjects the difference between these two conditions takes longer time to develop. Such individual differences can be seen in Figure 6a, where the differences in amplitude between the F2 and B2 conditions for electrodes included in the pre-defined ROI are drawn for each subject, for every 200 msec bin from 200 msec to 1000 msec after memory array onset. It can be seen that the difference (F2-B2) was positive in most of the subjects and most of the time bins, although subjects varied as to the latency of the maximum difference. Therefore, we extended the temporal ROI into the whole maintenance period from 200 msec to 1000 msec after the array onset and segregated it into four 200 msec segments. We then averaged the amplitudes across the ROI channels for each segment. For each subject, we then selected the temporal bin that elicited the largest difference between F2 and B2, and compared the amplitudes of the F2 and B2 conditions across subjects with paired sample t-test. We compared the result to a distribution under the null hypothesis generated through a permutation procedure. In each iteration of this procedure we randomly permuted the labels of the B2 and F2 conditions and then applied the same procedure of segmentation and peak difference selection to compare the surrogate conditions with a paired sample t-test. Thus, each permutation contributed one value to the null distribution and this procedure was repeated 5000 times. The t-value of the empirical data (red vertical line in Figure 6b) was larger than 99.48 % of t-values in the permutated trials (Histogram in Figure 6b),

which is unlikely to be obtained by chance. Thus, allowing temporal variance in the ROI of each individual leads to replication of the effect found in experiment 1, showing that the F2 condition elicits larger positivity compared with B2 condition during the maintenance duration. Critically, we repeated the procedure, now looking for the maximum negative (F2<B2) difference between F2 and B2, but the result was far from significant in this case (Figure 6c).

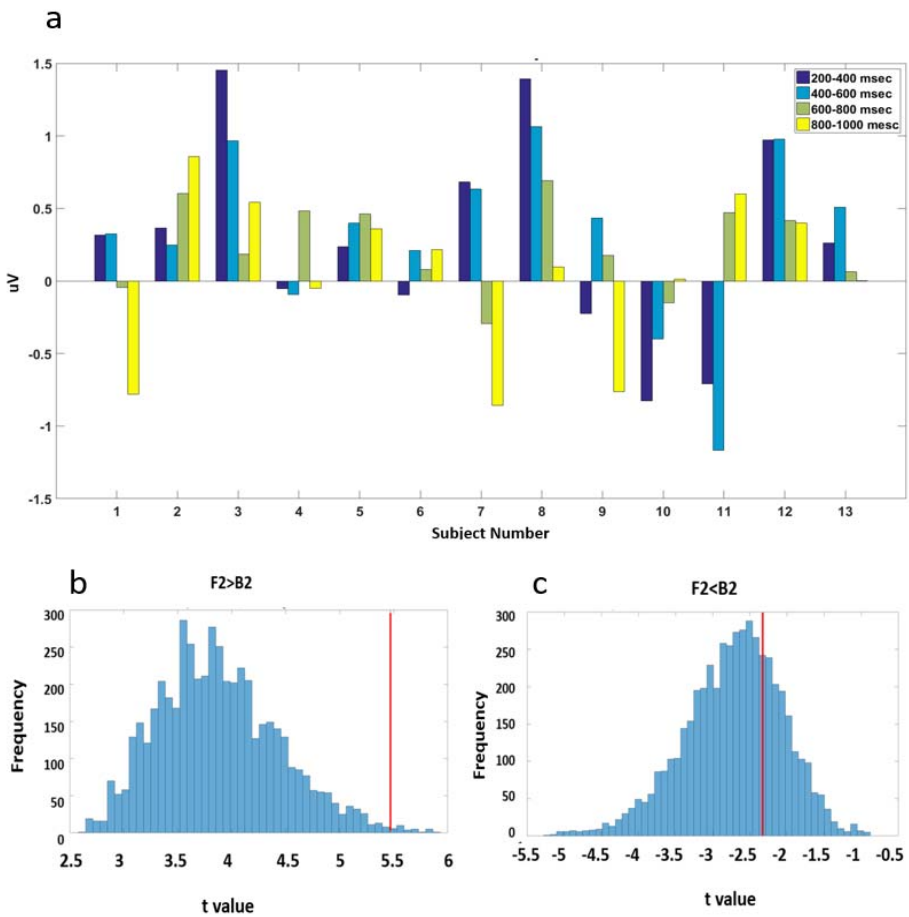

Figure 6 *: **a:** Amplitude differences between F2 condition and B2 condition (F2-B2) averaged across pre-defined spatial ROI for each subject within each 200 msec time window covering from 200 msec to 1000 msec after the memory array onset. **b**: The histograms, built by 5000 times randomly shuffling the label of B2 and F2 condition, represent the null distribution of t-values. The red lines represent the t-value of the empirical data. The t-value of maximum positive differences between F2 and B2 from 200 msec to1000 mesc after the onset of memory array is larger than 99.48 % results in null distribution. **c**: The t-value of maximum negative differences between F2 and B2 from 200 msec to1000 mesc after the onset of memory array is only smaller than 28.94% of the null distribution.*

MVPA:

If conditions in Experiment 2 could be successfully "decoded" from the classifier trained in Experiment 1, it would also support the notion that the same binding effect

was found across the two experiments. Therefore, we trained group-based classifiers at each time point for experiment 1 by pooling over all subjects and averaging each random 5 trials (sampled without replacement[4]) for each condition. The resulting classifiers were then applied to decode B2 from F2 condition for each subject in experiment 2 at each time point. The testing set for each subjects consisted of averages of 5 trials randomly selected with replacement. Classifiers trained in Experiment 1 were tested on every time point ranging from 200 msec before the stimuli onset until 1100 after the stimuli onset of each subjects in Experiment 2, resulting in a temporal generalization matrix of AUC for each subject. These results were then compared with baseline (*AUC=0.5*) at the group level by cluster based permutation test as used in the previous analysis. Classifiers trained during 310 msec to 538 msec after the stimuli onset in Experiment 1 could decode successfully conditions in the Experiment 2 from 314 msec to 851 msec (*P=0.049 clustered corrected*, *Figure 7*). This result is consistent with the observation that the same "binding effect" indicated by the contrast between B2 and F2 condition in the first group of subjects spread into later times in the second group of subjects. Regardless of temporal window, this cross-experiment decoding analysis suggests that the topographical distribution characterizing the "binding effects" found in Experiment 1 was successfully replicated in Experiment 2.

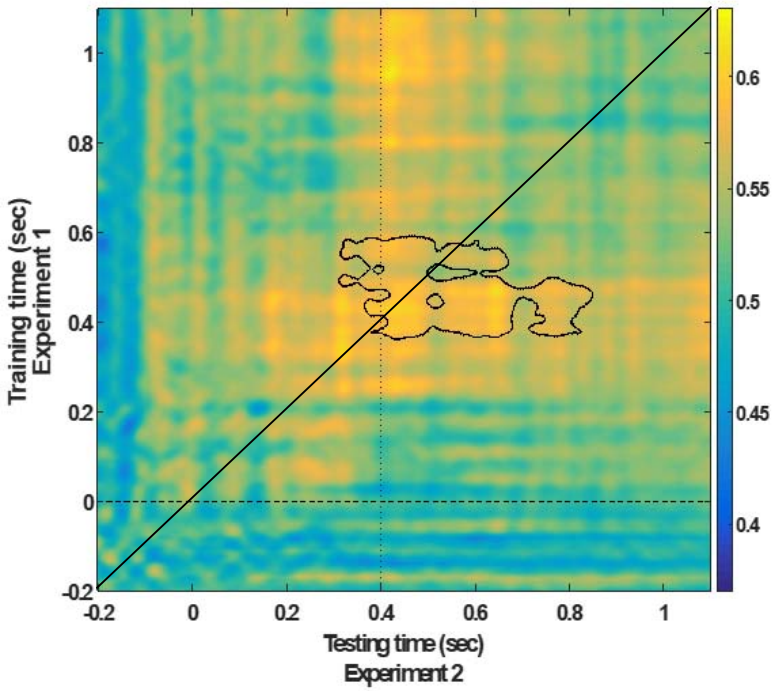

Figure 7: *Temporal generalization metrics that train on experiment 1 and test on experiment 2. Training time is shown on the vertical axis and testing time on the horizontal axis. Significant clusters are indicated by black contours.*

[4] Unlike the single subject decoding, the large number of trials in the pooled set made sampling with replacement unnecessary and computationally tasking.

Discussion

In two experiments, we found significant differences in EEG signals between tasks requiring short term maintenance of visual features and maintenance of bound objects, even when the visual input was identical across tasks. This provides evidence of task-dependent maintenance processes and against the notion, implicitly embedded in previous literature, that a fixed default mode of item representation is maintained in visual working memory. Furthermore, our results suggest that maintaining information in the bound-object condition may require less processing resources than in the separate-feature condition.

As expected, the ERP results reflected stronger central positivity in the more demanding B4 condition, compared to the B2 condition. Indeed, the behavioral results indicated that the B4 condition was more challenging than the other conditions. Considering the topographical distribution of this difference and the temporal dynamic, it is reasonable to take this increased positivity as a marker of memory load, but we cannot rule out a pure visual effect as the number of items on the display differed between B4 and the other conditions. Conversely, the B2 and F2 conditions consisted of identical arrays of two items, and lead to similar report accuracy. Still, a significant difference between the ERPs in these conditions was observed. Thus, the difference between the F2 and the B2 conditions was task-related rather than stimulus-related. The F2 condition elicited a higher central positivity and was more akin to the B4 (high load) condition than to the B2 condition in that respect. This pattern could reflect either (a) a larger number of units (i.e. 4 feature values vs. 2 bound objects) maintained during the delay period of F2 vs. B2 or (b) extra activity in the F2 condition when bound objects had to be decomposed into separate features to efficiently fulfill the task requirement. We note a more frontal topography for the B2 and B4 difference relative to the B2 and F2 difference, which might suggest different processes, favoring option (b) above. This is also confirmed by across decoding results, suggesting that F2 and B2 contrast does not reflect mainly the putative load effect. We also note that the MVPA analysis could discriminate F2 from B2 for a relatively limited duration of around 400 ms, which may fit more with a time-limited operation than with ongoing maintenance. However, these two explanations are not mutually exclusive and cannot be completely differentiated in the current study.

A previous visual search study (Berggren & Eimer, 2018) provided some intuition on how the brain represents information in a format that is compatible with the test demands. In this experiment, subjects answered whether the search display contained one of two target items held in working memory. When the search array included only one object with a target matching feature, targets and incorrect conjunction objects elicited identical N2pc component and sustained posterior contralateral negativity (SPCN). The N2pc is assumed to index a shift of spatial attention to the location of the potential target (Eimer, 1996; Kiss, Van , & Eimer, 2008), and the SPCN is assumed to index attentional activation of visual working memory representations of the potential target (Jolicoeur, Brisson, & Robitaille, 2008). These results suggest that in this condition only features were used as searching templates. In another condition, it was insufficient to detect a certain feature, as all objects had target-matching features, and both a target and an incorrect conjunction object could be present in the

same display. In this case, the target evoked a larger N2pc than the incorrect conjunction objects, and only targets elicited SPCN components. Therefore, in this condition a bound object, rather than an isolated feature template, was used. Taken together, the results of this study are compatible with the possibility that working memory templates guiding attention in visual search were encoded or used flexibly, to effectively distinguish match and no-match probes, based on the task demand.

In our study, a color probe in the feature condition (F2) always appeared in the center (a location never used in the memory arrays) and a location probe was always colored black (a color never used in the memory arrays). Thus, the matched-probes in F2 (feature) condition matched only one feature dimension of the stimuli in the memory array, whereas the other feature of the probe was fixed across trials, with a value that never appeared in the memory array. In this condition, it is more efficient to maintain feature values which are not yoked to other feature dimensions. Since the memory array contained two colors and two locations, 4 different feature values had to be maintained. The matched-probes in the B2 (binding) condition, on the other hand, matched an item from the memory array in both feature dimensions. In this case, retaining bound objects allows a direct comparison between the representation in the working memory and the upcoming probe, and may lead to more efficient performance.

Task-specific encoding and retention of visual stimuli have been observed by other studies addressing working memory more directly. For example, one study found that memory for different features can be differentially affected by retro-cues indicating which feature dimension would be tested (Park, Sy, Hong, & Tong, 2017). Woodman and Vogel (2008) suggested that object colors and orientations could be represented in working memory either separately, or together, according to the task requirement. A lower contralateral delay activity (CDA), a marker of memory load, was observed when only color was task-relevant compared to the case both color and orientation were relevant. Using CDA, a recent study (McCants, Katus, & Eimer, 2020) showed that objects were represented separately by their parts when parts appeared as probes, whereas a single compound object was maintained when objects were used as probes. McCants et al. concluded that not only can top-down control determine which of several feature dimensions (e.g. color, orientation) to maintain, but also whether parts of objects or their gestalt would be maintained. Here, we showed that the system is even more flexible – it allows selection between the maintenance of separate feature dimensions vs. a bound object according to the task goal and the nature of the probe.

The finding that objects are retained either as separate features or as bound items based on task requirements may explain some of the seemingly conflicting results in the literature, although not all of them. This is based on the idea that although two tasks are similar in their explicit instructions, subtle differences in the relationship between the memory array and the probe might affect the way that the memory array is maintained in memory. For example, Luck and Vogel (1997; extended in Vogel, Woodman, & Luck, 2001) observed that the ability to detect a change in a stimulus array of a given set size is the same regardless of the number of features. In contrast, Wheeler and Treisman (2002) found memory capacity is determined merely by the number of total features when two features from the dimension were combined into

one object. In Luck and Vogel (1997), each item presented in the memory array contained features randomly selected from all possible feature values with replacement (e.g., more than one item could be red) and the non-match probe could include an erroneous conjunction of features that appeared in the memory array. In contrast, in the study by Wheeler and Treisman (2002, Experiments 1 and 2), items on the memory array were generated by selecting from possible feature values without replacement (e.g. only one item could be red). Moreover, to generate a non-matched probe, the probe included a feature that had not been used by any item in the memory array. This design, where no replacement was allowed when assigning feature value to a single memory array, can be also found on other studies claiming that increased features to be remembered for each object impaired change detection (Oberauer & Eichenberger, 2013). In the study by Luck and Vogel (1997), bound objects had to be maintained to detect a non-match probe, while in studies showing contradictory observations, remembering a list of all features in the memory array was sufficient to distinguish between matched and non-match probe. The different results regarding representation format might therefore reflect the fact that such format is flexibly shaped by task goal. The current findings, therefore, draw attention to subtle details between experimental designs which could have an impact on the pattern of information that is maintained in memory.

Whereas the electrophysiological difference between the B2 and B4 condition was accompanied by a difference in task accuracy, the same difference between B2 and F2 conditions did not result in an observable change in accuracy. This highlights a dissociation between the activity measured continuously during the maintenance period, and the decision recorded after the presentation of the probe. Taking the results together, one can speculate that more units held in working memory are reflected by enhanced positivity (and possibly an enhancement of the CDA when stimuli are lateralized), whereas diminished accuracy reflects increase in difficulty of comparing the probe with what was retained in working memory. For example, Alvarez and Cavanagh (2007) showed that working memory capacity is inversely correlated with item complexity. However, Awh et al (2007) showed that the sample-test similarity is highly correlated with item complexity, and that the item complexity effect on memory performance was not observed when the sample-test similarity was low. This suggests that the retrieval process, rather than the maintenance stage, limits memory performance. Further studies could test this hypothesis more directly.

It is worth noting that we have used location-to-color binding in our study, and this kind of binding has some unique aspects that might not generalize to other types of binding. There is some evidence suggesting that objects are automatically encoded together with location. For example, some studies showed the task-irrelevant location of a stimulus can also be directly decoded from EEG data during the delay period of a working memory task (Elsley & Parmentier, 2015; Foster, Bsales, Jaffe, & Awh, 2017; Olson & Marshuetz, 2005). EEG studies show that spatial attention can be drawn to items in WM even when the cue is non-spatial and the location is entirely irrelevant for the task (Eimer & Kiss, 2010; Kuo, Rao, Lepsien, & Nobre, 2009). There are studies supporting an even stronger claim that non-spatial features binding is mediated by binding to locations (Schneegans & Bays, 2017; Treisman & Gelade,

1980; Pertzov & Husain, 2014), perhaps due to the fact that most neurons selective to visual features also exhibit some spatial selectivity (Schneegans and Bays, 2019). Further studies are needed to test whether the results in the current test generalize to cases in which objects are defined by two non-spatial features.

To conclude, our study supports the view that working memory should not be construed as a passive storage of environmental input, but a dynamic system actively engaging in the task and preparing to carry out an effective response. Specifically, it provides a novel piece of evidence that the format of objects represented in the working memory, either as bound objects that integrate different feature dimensions or as separated features, is task-dependent. The finding of flexible maintenance of features and bound objects calls for deeper consideration of how experimental manipulations, either implicitly or explicitly, influence the format of objects maintained in working memory. Additionally, although the underlying mechanism demands further investigation, throughout two studies we found evidence to reject an intuitive suggestion that maintaining bound objects is more resource-demanding than maintaining the same number of features included in that object.

Declaration of interests

Ruoyi Cao, Yoni Pertzov, Mowei Shen and Zaifeng Gao declare no competing interests. Leon Y. Deouell is a co-founder, advisor and equity holder of Innereye Ltd which has no direct interest in the current study.

Acknowledgements

The study was supported by a China-Israel cooperative *scientific* research provided by the Israel ministry of Science, Technology and Space to Leon Y. Deouell and Yoni Pertzov, and Ministry of Science and Technology of the People's Republic of China (2016YFE0130400).to Mowei Shen. We thank Yongdi Zhou for helping with design and discussion of the present study. We thank Naomi Revel and Geffen Markusfeld for helping with running experiments.

References

Alvarez, G. A., & Cavanagh, P. (2004). The Capacity of Visual Short-Term Memory Is Set Both by Visual Information Load and by Number of Objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628. https://doi.org/10.1111/j.1467-9280.2007.01949.x

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory□: The role of the episodic buffer. *Neuropsychologia*, *49*(6), 1393–1400. https://doi.org/10.1016/j.neuropsychologia.2010.12.042

Baizer, J. S., Ungerleider, L. G., & Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *Journal of Neuroscience*, *11*(1), 168–190. https://doi.org/10.1523/jneurosci.11-01-00168.1991

Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1996). Object and Spatial Visual Working Memory Activate Separate Neural Systems in Human Cortex. *Cerebral Cortex*, *6*(1), 39–49. https://doi.org/10.1093/cercor/6.1.39

D'Esposito, M., Aguirre, G. K., Zarahn, E., Ballard, D., Shin, R. K., & Lease, J. (1998). Functional MRI studies of spatial and nonspatial working memory. *Cognitive Brain Research*, *7*(1), 1–13. https://doi.org/10.1016/S0926-6410(98)00004-4

Delvenne, J. F., & Bruyer, R. (2004). Does visual short-term memory store bound features? *Visual Cognition*, *11*(1), 1–27. https://doi.org/10.1080/13506280344000167

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, *99*(3), 225–234. https://doi.org/10.1016/s0921-884x(96)95711-2

Eimer, Martin, & Kiss, M. (2010). An electrophysiological measure of access to representations in visual working memory. *Psychophysiology*, *47*(1), 197–200. https://doi.org/10.1111/j.1469-8986.2009.00879.x

Elsley, J. V., & Parmentier, F. B. R. (2015). The asymmetry and temporal dynamics of incidental letter–location bindings in working memory. *Quarterly Journal of Experimental Psychology*, *68*(3), 433–441. https://doi.org/10.1080/17470218.2014.982137

Foster, J. J., Bsales, E. M., Jaffe, R. J., & Awh, E. (2017). Alpha-Band Activity Reveals Spontaneous Representations of Spatial Position in Visual Working Memory. *Current Biology*, *27*(20), 3216-3223.e6. https://doi.org/10.1016/j.cub.2017.09.031

Fougnie, D., & Marois, R. (2009). Attentive tracking disrupts feature binding in visual working memory. *Visual Cognition*, *17*(1–2), 48–66. https://doi.org/10.1080/13506280802281337

Hardman, K. O., & Cowan, N. (2015). Remembering complex objects in visual

working memory: Do capacity limits restrict objects or features? *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*(2), 325–347. https://doi.org/10.1037/xlm0000031

He, K., Li, J., Wu, F., Wan, X., Gao, Z., & Shen, M. (2020). Object-based attention in retaining binding in working memory: Influence of activation states of working memory. *Memory and Cognition*, *48*(6), 957–971. https://doi.org/10.3758/s13421-020-01038-0

Jolicœur, P., Brisson, B., & Robitaille, N. (2008). Dissociation of the N2pc and sustained posterior contralateral negativity in a choice response task. *Brain Research*, *1215*, 160–172. https://doi.org/10.1016/j.brainres.2008.03.059

Kiss, M., Van Velzen, J., & Eimer, M. (2008). The N2pc component and its links to attention shifts and spatially selective visual processing. *Psychophysiology*, *45*(2), 240–249. https://doi.org/10.1111/j.1469-8986.2007.00611.x

Kravitz, D. J., Kriegeskorte, N., & Baker, C. I. (2010). High-level visual object representations are constrained by position. *Cerebral Cortex*, *20*(12), 2916–2925. https://doi.org/10.1093/cercor/bhq042

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, *12*(4), 217–230. https://doi.org/10.1038/nrn3008

Kuo, B. C., Rao, A., Lepsien, J., & Nobre, A. C. (2009). Searching for targets within the spatial layout of visual short-term memory. *Journal of Neuroscience*, *29*(25), 8032–8038. https://doi.org/10.1523/JNEUROSCI.0952-09.2009

Liang, Y., Pertzov, Y., Nicholas, J. M., Henley, S. M. D., Crutch, S., Woodward, F., … Husain, M. (2016). ScienceDirect Visual short-term memory binding deficit in familial Alzheimer ' s disease. *CORTEX*, *78*, 150–164. https://doi.org/10.1016/j.cortex.2016.01.015

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(1996), 279–281.

Luria, R., & Vogel, E. K. (2011). Shape and color conjunction stimuli are represented as bound objects in visual working memory. *Neuropsychologia*, *49*(6), 1632–1639. https://doi.org/10.1016/j.neuropsychologia.2010.11.031

Makovski, T., & Pertzov, Y. (2015). Rapid communication Attention and memory protection□: Interactions between retrospective attention cueing and interference, *68*(9), 1735–1743. https://doi.org/10.1080/17470218.2015.1049623

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

McCants, C. W., Katus, T., & Eimer, M. (2020). Task goals modulate the activation of part-based versus object-based representations in visual working memory. *Cognitive Neuroscience*, *11*(1–2), 92–100. https://doi.org/10.1080/17588928.2019.1642864

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, *6*(C), 414–417.

https://doi.org/10.1016/0166-2236(83)90190-X

Morey, C. C., & Bieler, M. (2013). Visual short-term memory always requires general attention. *Psychonomic Bulletin and Review*, *20*(1), 163–170. https://doi.org/10.3758/s13423-012-0313-z

Myers, N. E., Stokes, M. G., & Nobre, A. C. (2017). Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends in Cognitive Sciences*, *21*(6), 449–461. https://doi.org/10.1016/j.tics.2017.03.010

Nobre, A. C. (2007). Attentional Modulation of Object Representations in Working Memory, (September). https://doi.org/10.1093/cercor/bhl116

Oberauer, K., & Eichenberger, S. (2013). Visual working memory declines when more features must be remembered for each object. *Memory and Cognition*, *41*(8), 1212–1227. https://doi.org/10.3758/s13421-013-0333-6

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Perception and Psychophysics*, *64*(7), 1055–1067. https://doi.org/10.3758/BF03194756

Olson, I. R., & Marshuetz, C. (2005). Remembering "what" brings along "where" in visual working memory. *Perception and Psychophysics*, *67*(2), 185–194. https://doi.org/10.3758/BF03206483

Park, Y. E., Sy, J. L., Hong, S. W., & Tong, F. (2017). Reprioritization of Features of Multidimensional Objects Stored in Visual Working Memory. *Psychological Science*, *28*(12), 1773–1785. https://doi.org/10.1177/0956797617719949

Parra, M. A., Abrahams, S., Fabi, K., Logie, R., Luzzi, S., & Sala, S. Della. (2009). Short-term memory binding deficits in Alzheimers disease. *Brain*, *132*(4), 1057–1066. https://doi.org/10.1093/brain/awp036

Parra, M. A., Della Sala, S., Logie, R. H., & Morcom, A. M. (2014). Neural correlates of shape-color binding in visual working memory. *Neuropsychologia*, *52*(1), 27–36. https://doi.org/10.1016/j.neuropsychologia.2013.09.036

Pertzov, Y., Miller, T. D., Gorgoraptis, N., Caine, D., Schott, J. M., Butler, C., & Husain, M. (2013). Binding deficits in memory following medial voltage-gated potassium channel complex antibody-associated limbic encephalitis, 2474–2485. https://doi.org/10.1093/brain/awt129

Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science*, *276*(5313), 821–824. https://doi.org/10.1126/science.276.5313.821

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), 1–8. https://doi.org/10.1111/psyp.13335

Schneegans, S., & Bays, P. M. (2017). Neural architecture for feature binding in visual working memory. *Journal of Neuroscience*, *37*(14), 3913–3925. https://doi.org/10.1523/JNEUROSCI.3493-16.2017

Schneegans, S., & Bays, P. M. (2019a). New perspectives on binding in visual working memory. *British Journal of Psychology*, *110*(2), 207–244.

https://doi.org/10.1111/bjop.12345

Schneegans, S., & Bays, P. M. (2019b). New perspectives on binding in visual working memory. *British Journal of Psychology*, *110*(2), 207–244. https://doi.org/10.1111/bjop.12345

Shen, M., Huang, X., & Gao, Z. (2015). Object-based attention underlies the rehearsal of feature binding in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 479–493. https://doi.org/10.1037/xhp0000018

Smith, E. E., Jonides, J., Koeppe, R. A., Awh, E., Schumacher, E. H., & Minoshima, S. (1995). Spatial versus object working memory: PET investigations. *Journal of Cognitive Neuroscience*, *7*(3), 337–356. https://doi.org/10.1162/jocn.1995.7.3.337

Treccani, B. (2018). The neuropsychology of feature binding and conscious perception. *Frontiers in psychology*, 9, 2606. 1–5. https://doi.org/10.3389/fpsyg.2018.02606

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London.* Series B: Biological Sciences, 353(1373), 1295-1306.

Treisman, A. M., & Garry Gelade. (1980). A feature-integration theory of attention. *Cognitive psychology, 12(1), 97-136.*

Ueno, T., Allen, R. J., Baddeley, A. D., Hitch, G. J., & Saito, S. (2011). Disruption of visual feature binding in working memory. *Memory and Cognition*, *39*(1), 12–23. https://doi.org/10.3758/s13421-010-0013-8

Vergauwe, E., Langerock, N., & Barrouillet, P. (2014). Maintaining information in visual working memory: Memory for bindings and memory for features are equally disrupted by increased attentional demands. *Canadian Journal of Experimental Psychology*, *68*(3), 158–162. https://doi.org/10.1037/cep0000025

Wang, B. (2017). Separate Capacities for Storing Different Features in Visual Working Memory Journal of Experimental Psychology□: Learning , Memory , and Cognition, *43*(July), 226–236. https://doi.org/10.1037/xlm0000295

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*(1), 48–64. https://doi.org/10.1037/0096-3445.131.1.48

Woodman, G. F., & Vogel, E. K. (2008). Selective storage and maintenance of an object's features in visual working memory. *Psychonomic Bulletin and Review*, *15*(1), 223–229. https://doi.org/10.3758/PBR.15.1.223