The time-course of component processes of selective attention

Tanya Wen[1]*, John Duncan[1,2], and Daniel J Mitchell[1]

[1]Medical Research Council, Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 7EF, United Kingdom

[2]Department of Psychology, University of Oxford, Oxford, UK

Abbreviated title: Time-course of selective attention processes

*** Corresponding author:** email address: tanya.wen@mrc-cbu.cam.ac.uk

Number of pages: 38

Number of figures: 7

Number of Tables: 1

Number of words: Abstract = 239, Introduction = 649, Discussion = 1437

Conflict of Interest: The authors declare no competing financial interests.

Acknowledgements: This work was supported by funding from the Medical Research Council (United Kingdom), program SUAG/002/RG91365. TW was supported by the Taiwan Cambridge Scholarship from the Cambridge Commonwealth, European & International Trust and the Percy Lander studentship from Downing College.

Abstract

Attentional selection shapes human perception, enhancing relevant information, according to behavioral goals. Here, we used multivariate decoding of electrophysiological brain responses (MEG/EEG) to examine real-time representation of the component processes of selective attention. Auditory cues instructed participants to respond to a particular visual target, embedded within a stream of single- and multi-item displays. Although the task logically required items to be compared to an attentional "template", signals consistent with such a template were relatively weak and appeared to transition through a quiescent state. Subsequent stimulus processing evoked strong neural representation of multiple target features, evolving over different timescales. Combining single and multi-item displays with different types of distractors allowed quantification of various components of attention. Following a visual choice display, we observed five distinguishable processing operations with different time-courses. First, visual properties of the stimulus were strongly represented. Second, the location of the candidate target was rapidly represented in multi-item displays, providing the earliest evidence of modulation by behavioral relevance. Third, there was enhanced representation of the candidate target, including its identity, relative to distractors. Fourth, only later was the behavioral significance of the target explicitly represented in single-item displays. Finally, if the target was not identified and search was to be resumed, then an attentional template was weakly reactivated. The observation that an item's behavioral relevance directs attention in multi-item displays prior to explicit representation of target/non-target status in single-item displays is consistent with two-stage models of attention.

Significance Statement

Imagine looking for a friend in a crowd. Visual search starts from an internal template of the target, and attentional selection enhances representation of information consistent with the target. This study examined the component processes of attentional selection using MEG/EEG. Preparatory representation of an upcoming target was weak. The choice display then evoked distinct representations of visual features, target location, target identity, and behavioral significance, each with a characteristic time-course. Intriguingly, evidence of directed attention to a target appeared earlier than explicit categorization of a single stimulus. Multiple processes of selective visual attention can be robustly decoded from MEG/EEG signals.

Introduction

Our perception of the world is constantly shaped by attentional selection, enhancing relevant over irrelevant information, to achieve our behavioral goals. Selection must begin from a flexible description, often called the attentional template, of the object currently required (Duncan and Humphreys, 1989; Bundesen, 1990). Much evidence suggests that attentional selection is achieved through a process of biased, integrated competition across a broad sensorimotor network (Duncan et al., 1997). As objects in the visual input compete to dominate neural activity, the degree to which they match the attentional template determines their competitive advantage (Desimone and Duncan, 1995; Beck and Kastner, 2009).

Recently, there has been much interest in the use of MEG/EEG for real-time decoding of cognitive events in the human brain (Stokes et al., 2015). Here, we used simultaneous MEG/EEG to examine the time-course and content of attentional selection. The behavioral relevance of stimuli was manipulated by starting each trial with one of two auditory cues,

indicating the relevant target object on this trial. Participants were then presented with a series of visual displays of 4 possible types: a 1-item display of the target (T), an inconsistent non-target (Ni; which was associated with the other cue and served as a target for other trials), a consistent non-target (Nc; which was never a target), or a 3-item display with all items presented simultaneously. Participants were required to make a button press whenever they detected a brightening of the target item. Using multivariate decoding analyses, we asked which component processes of attentional selection are visible in the MEG/EEG signal over time.

First, we examined representation of the attentional template. One possibility is that, when a cue indicates the relevant target object, some sustained signal will be set up in neurons selectively responsive to that object (Chelazzi et al., 1993; Puri et al., 2009; Kok et al., 2013). fMRI decoding studies have shown cross-generalization between attentional templates and sensory responses to the corresponding objects (e.g., Stokes et al., 2009; Peelen and Kastner, 2011), supporting a tonic activation of visual representations. However, corresponding results tend to be weak or non-existent in electrophysiological recordings (Stokes et al., 2013; Myers et al., 2015; Wolff et al., 2015), and where they have been found, they may appear only very briefly prior to the target stimulus (Myers et al., 2015; Kok et al., 2017). Indirect measures of attentional templates, derived from ERP components, demonstrated that search templates are not continuously active but are transiently activated during the preparation for each new search episode (Grubert and Eimer, 2018). Recently, it has been proposed that template storage may often be "silent", perhaps encoded in changed synaptic weights rather than sustained firing (Stokes, 2015). To examine template coding in MEG/EEG, holding visual input constant, we analyzed data from the period between cue and displays, and during subsequent presentation of Nc stimuli.

Second, we were interested in the process of target detection itself. Comparing target and non-target stimuli shows strong differences both behaviorally and neurally (Duncan, 1980; Hebart et al., 2018). Attending to a relevant visual object produces strong, sustained activity across many brain regions (Desimone and Duncan, 1995; Sergent et al., 2005; Dehaene and Changeux, 2011), reflecting encoding of its multiple visual properties and implications for behavior (Wutz et al., 2018). In the presence of multiple stimuli, neural responses are initially divided amongst the competing sensory inputs and later become replaced by a wide-spread processing of the behaviorally critical target (Duncan et al., 1997; Kadohisa et al., 2013). On 1-item trials, we focused on the response to the T and Ni stimuli, to quantify the representation of object identity (e.g., face vs. house) regardless of status as target or non-target, as well as representation of behavioral category (T vs. Ni) regardless of object identity. On 3-item trials, we quantified the encoding of target location and target identity, to assess preferential processing of target features when multiple items compete for representation.

Methods

Participants

Eighteen participants (9 males, 9 females; age range: 18-30 years, mean = 24.4, SD = 3.8) took part in the study, recruited from the volunteer panel of the MRC Cognition and Brain Sciences Unit. Two additional participants were excluded from the analysis due to technical problems (one could not do the MRI; another was excluded due to an error in digitizing the EEG electrodes). EEG data for 4 participants were excluded from the MVPA analysis due to a technical issue (a test signal used during hardware checkup was not removed). All

participants were neurologically healthy, right-handed, with normal hearing and normal or corrected-to-normal vision. Procedures were carried out in accordance with ethical approval obtained from the Cambridge Psychology Research Ethics Committee, and participants provided written, informed consent prior to the experiment.

Stimuli and Procedures

Participants performed two localizer tasks (auditory and visual) and an attention task (see Figure 1 for an illustration). Stimulus presentation was controlled using the Psychophysics Toolbox (Brainard, 1997) in Matlab 2014a (Mathworks, Natick, WA). Auditory stimuli were delivered through in-ear headphones compatible with the MEG recording. Visual stimuli were back-projected onto a screen placed in front of the participant, approximately 129 cm from the participant's eyes. Each stimulus image was approximately 20 cm wide (visual angle 8.8°) on a gray background. Before the start of each task, participants were given training to familiarize them with the stimuli and task rules. If a false alarm was made during any of the trials during the recording, that trial was repeated at the end of the run.

Pattern Localizer Tasks

Auditory Localizer Task: This task was used to characterize multivariate activity patterns for high and low pitch tones used in the attention task. Participants heard a stream of intermixed high (1100 Hz) and low (220 Hz) pitch tones. On rare occasions (9% of the time), a frequency modulation would occur (modulator frequency = 20 Hz; modulation index = 2.5), and participants were instructed to press a button whenever they detected a distortion in a tone. There were 100 occurrences of each unmodulated tone and 10 occurrences of each modulated tone. The duration of each tone was 100 ms, with the beginning and ending 10 ms ramped. The inter-stimulus interval was jittered between 1000-1500 ms.

Visual Localizer Task: Similar to the auditory localizer task, this task was used to establish multivariate activity patterns for three visual stimuli (a face, a house, and a violin) used in the attention task. Participants were shown a stream of these images presented sequentially in the center of the screen for 100 ms each, with an inter-stimulus interval jittered between 1500-2000 ms. Most image displays were semi-transparent (60% opaque) on a grey background; participants were asked to make a button press whenever they detected a brighter version of the image (100% opaque). There were 100 occurrences of each translucent image and 10 occurrences of each brightened image.

Attention Task

Figure 1 illustrates the stimuli used in the task, as well as the task structure. Before the start of this task, participants were trained to associate the two auditory tones with two of the three visual stimuli (the same used in the localizer tasks). This pairing resulted in the visual stimuli being categorized by behavioral relevance as targets (T: the visual stimulus paired with the current cue), inconsistent non-targets (Ni: the visual stimulus paired with the alternative cue), and consistent non-targets (Nc: never targets). The pairings were counterbalanced across participants. The task was executed in runs of 90 trials. Each trial began with an auditory cue (for 100 ms), followed by a 1750 – 2250 ms fixation cross during which participants were instructed to prepare to attend for the target stimulus. Then a stream of three visual displays appeared one by one for 100 ms each, separated by 1500-2000 ms inter-stimulus intervals. Each display could be a 1-item display or a 3-item display with equal probability (order pseudorandomized, with the constraint that a 1-item display could not follow a 1-item display of the same type). On 1-item displays, the stimulus was centered at fixation; 3-item displays contained all three visual stimuli, each 8.8° visual angle from fixation, arranged in an

equilateral triangle with one above left, one above right, and one below. In 18 out of the 90 trials in each run, a single brightened stimulus, target or non-target, occurred pseudo-randomly in one of the 3 displays, with equal likelihood of appearing in each. For each cue type, brightenings affected T, Ni and Nc items once each on single item-trials, and twice each on 3-item displays, allowing one brightening for each of the six possible 3-item stimulus configurations. Participants were asked to attend to targets, pressing a button if they detected a brightened target (they could respond any time before the next stimulus), with no response for all other displays. The trial terminated if a button press was made, and participants were informed whether the response was a correct detection or a false alarm. A new trial began when the participant indicated with a button press that they were ready to continue. Otherwise, each of the 90 trials in each run had a full sequence of 3 displays. At the end of each run, feedback informed participants of their accuracy through the run. To discourage false alarms and equalize the number of non-response trials across conditions, trials that contained a false alarm were repeated at the end of the run. The task was repeated over 5 runs (2 participants only completed 4 runs due to time constraints).

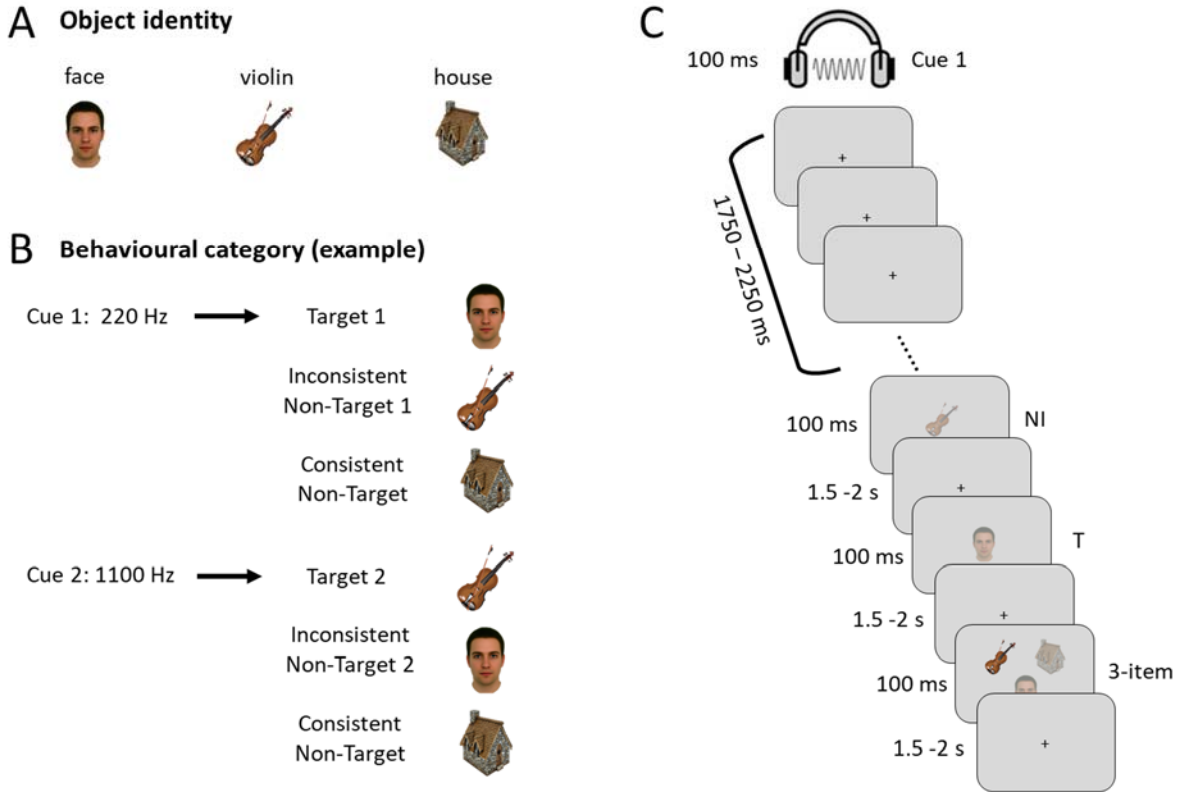


Figure 1. Stimuli and experimental paradigm. (A) The 3 objects used in the experiment. (B) An example of how the two auditory tones could be paired with the three objects. This results in two items that serve as targets (T) for one cue, and non-targets (Ni) for the other cue, and the third item serving as a consistent non-target (Nc). The pairings between the tones and the objects were counterbalanced across participants. (C) An example trial illustrating the experimental paradigm. At the beginning of each trial, an auditory cue indicated the target for that trial. After a delay, this was followed by three visual displays. Participants were asked to make an immediate button press if a brightening of the target stimulus was detected.

Data acquisition

Electroencephalography (EEG)

EEG data were collected from 70 Ag/AgCl electrodes mounted on an electrode cap (Easycap, Falk Minow Services, Herrsching-Breitbrunn, Germany) distributed according to the extended 10/20 system. Electrode impedances were kept below 5 kΩ. An electrode placed on

the nose served as online reference while the ground electrode was placed on the right cheek. Vertical and horizontal eye movements were monitored using the electrooculograms (EOG) recorded using bipolar electrodes placed above and below the left eye and at the outer canthi of the eyes, respectively. Electrocardiography (ECG) was recorded using bipolar electrodes placed below the right collarbone and below the left ribcage. EEG data were sampled at 1000 Hz with a band-pass filter of 0.1–333 Hz. EEG and MEG data were acquired simultaneously.

Magnetoencephalography (MEG)

MEG data were acquired using a 306 channel (204 planar gradiometers and 102 magnetometers) Neuromag Vectorview system (Elekta AB, Stockholm) in a sound-attenuated and magnetically shielded room. Data were sampled at 1000 Hz with an online band-pass filter of 0.03–333 Hz. Five Head Position Indicator (HPI) coils were attached firmly to the EEG cap to track the head movements of the participant. The locations of the HPI coils as well as the EEG electrodes were recorded with a Polhemus 3D digitizer. We also measured three anatomical landmark points (nasion, left and right preauricular points) and additional points on the head to indicate head shape and enable matching to each individual's structural MRI scan.

Structural MRIs

High-resolution anatomical T1-weighted images were acquired for each participant (either after the MEG session or at least three days prior to the MEG session) in a 3T Siemens Tim Trio scanner, using a 3D MPRAGE sequence (192 axial slices, TR = 2250 ms, TI = 900 ms, TE = 2.99 ms, flip angle = 9°, field of view = 256 mm × 240 mm × 160 mm, 1 mm isotropic resolution). The coordinates of the nasion, left and right preauricular points in native space

were hand-marked by the experimenter, and used in the coregistration of the EEG/MEG and MRI.

EEG and MEG data preprocessing

The raw data were visually inspected during recording for any bad channels, which were removed (EEG: 0 - 5 across subjects; MEG: 1 - 5 across subjects). The MEG data were de-noised using Maxfilter 2.2 (Elekta Neuromag, Helsinki), with the spherical harmonic model centered on a sphere fit to the digitized head points; default settings were used for the number of basis functions and the spatiotemporal extension (Taulu & Simola, 2006). Maxfilter detected additional bad channels using the first and last 900 data samples (default threshold), and signal from all bad channels was removed and interpolated. Continuous movement compensation was applied at the same time.

Subsequent preprocessing used SPM12 (http://www.fil.ion.ucl.ac.uk/spm) and Matlab 2015a (The Mathworks Inc). Separately for EEG electrodes, magnetometers and gradiometers, independent component analysis (ICA), implemented using EEGLAB (Delorme & Makeig, 2004), was used to detect and remove components whose time-course correlated with EOG or ECG reference time-courses, and whose topography matched reference topographies associated with ocular or cardiac artefacts estimated from independent data acquired on the same system. ICA used the default infomax algorithm, with dimension reduction to 60 principal components. An independent component was removed if (1) it had the maximum absolute correlation with both a temporal and spatial reference, (2) these correlations were significant at $p < 0.05$, (3) the z-scored absolute correlations exceeded 2 for the spatial component, and 3 for the temporal component, and (4) it explained > 1.7% of total variance. For assessing temporal correlations only, ICA and reference time-courses were bandpass

filtered between 0.1 - 25 Hz, and correlations were also repeated 1000 times with phase randomization of the reference time-course to ensure that the true maximum absolute correlation of eliminated components was greater than the 95th percentile of the null distribution. EEG data were then re-referenced to the average reference.

Data were band-pass filtered between 0.1 Hz and 40 Hz and then epoched around the events of interest, time-locked to stimulus onset (from -100 ms to 1000 ms in the auditory localizer task; from -100 ms to 1500 ms in the visual localizer task; from -100 ms to 1750 ms for the cue and delay period of the main task, and -100 ms to 1500 ms for each of the visual stimulus presentations in the main task). Time points -100 ms to 0 ms served as baseline for baseline correction – the mean signal across this window was subtracted from each time point, per epoch. Epochs that contained flat segments or high threshold artifacts (peak-to-peak amplitude greater than 4000 fT for magnetometers, 400 fT/m for gradiometers, 120 μV for EEG, or 750 μV for EOG) were marked as bad trials and were rejected. In both localizer and attention tasks, any epoch that contained an auditory frequency distortion, a visual brightening, or a button press were additionally excluded from analyses. In the attention task, we also removed all data from any trial with an error (false alarm or miss). The average number of epochs remaining for each condition is shown in Table 1.

Table 1. Mean number of epochs (and standard deviation across participants) per condition after artifact rejection

Localizer Tasks	
Auditory localizer	
Low tone	High tone

80.8	81.0		
(16.8)	(14.9)		
Visual localizer			
Face	House	Violin	
72.1	70.1	69.7	
(13.6)	(14.9)	(15.4)	
	Attention Task		
Preparatory phase			
Low tone	High tone		
147.6	147.4		
(35.7)	(37.5)		
Stimulus processing phase			
Face	House	Violin	3-item
138.4	134.2	139.5	406.1
(36.4)	(34.2)	(34.3)	(110.2)
Target (T)	Inconsistent Non-target (Ni)	Consistent Non-target (Nc)	
138.9	137.9	135.1	
(35.3)	(35.0)	(34.8)	

Source Localization

For each participant, a cortical mesh was created from the individual's structural MRI, with a mesh resolution of ~4000 vertices per hemisphere. The EEG/MEG and MRI were

coregistered based on the three anatomical fiducial points and an additional ~200 digitized points on the scalp. Forward models were computed for EEG data using a three-shell boundary element model (BEM) and for MEG data using a single-shell BEM. The forward model was inverted using minimum norm estimation (MNE) to compute source estimates for each experimental condition. Due to the limited spatial resolution limits of EEG/MEG, we chose three *a priori* spatially distinct bilateral ROIs (Figure 2C). Early visual cortex and lateral prefrontal cortex (LPFC) were used to test representation in relevant sensory and cognitive control areas. An additional auditory cortex ROI was used both to measure cue decoding, and in other analyses to test for signal leakage. Auditory and primary visual cortex ROIs were taken from the SPM Anatomy toolbox (Eickhoff et al., 2005), containing 350 and 523 vertices. The LPFC ROI was taken from Fedorenko et al. (2013) (http://imaging.mrc-cbu.cam.ac.uk/imaging/MDsystem), combining the anterior, middle, and posterior medial frontal gyri, spanning 461 vertices.

Multivariate Pattern Analysis (MVPA)

Multivariate pattern analyses were performed using the Matlab interface of LIBSVM (Chang and Lin, 2011). We used a linear support vector machine, with default parameters. For each analysis, we performed decoding in sensor space as well as in source space using data from the three ROIs. For sensor space decoding, we combined data from good EEG and MEG (gradiometers and magnetometers) channels. Each individual time point was standardized (z-scored across channels) before entering the classifier. For source space decoding, each participant's cortical mesh was transformed into MNI space, and estimated source activity at each vertex within the ROIs was extracted to serve as a feature in the classifier.

In both sensor and source space MVPA analyses, we trained and tested on a spatiotemporal pattern extracted from a sliding time window of 32 ms. This was done every 4 ms. For analyses involving within-task decoding, the data were split into five folds (with one fold containing every 5th trial chronologically), iteratively trained on four of the folds and tested on the remaining fold. In cross-task decoding, a classifier was trained on all epochs from one task and tested on all epochs from another task. For within-task decoding at each time point, cross-validation accuracies were compared against chance (50%) with a one-tailed t-test, and multiple comparisons were accounted for using a false discovery rate (FDR) correction. Decoding outputs (classification accuracy) were temporally smoothed with a Gaussian filter for visualization (sigma = 20 ms), but statistical analyses were performed on unsmoothed data. For generalization across time or tasks, we first smoothed the (training) time x (testing) time matrices with a Gaussian filter (sigma = 20 ms), before performing one-tailed t-tests against chance and using cluster-based permutation tests to correct for multiple comparisons across time x time matrices (cluster-forming threshold: $p < 0.01$, number of permutations = 1000, significance criterion: $p < 0.05$; Myers et al. 2015). The figures were plotted according to the last time bin in the sliding window (Grootswagers et al., 2017). For decoding of 1-item behavioural category, epochs that were preceded by a T or Ni were excluded, to ensure that behavioural category was balanced in the baseline period.

Results

Behavioral results

Behavioral performance was consistently high (auditory localizer task – hits: mean = 99.0%; false alarms: mean = 0.8%; visual localizer task – hits: mean = 98.9%; false alarms: 1.0%; attention task – hits: mean = 98.3%; false alarms: mean = 0.8%).

Coding of the attentional cue/attentional template during the preparatory phase

Source localization of the response to the cue at representative time points is shown in figure 2A. We first looked for decoding of the specific attentional cue during the preparatory phase of the attention task, defined as starting from cue onset but before the first visual stimulus, and compared this with decoding in the auditory localizer task. We asked whether preparing for a target enhances cue decoding. Cue/stimulus decoding as a function of time from auditory stimulus onset is shown in Figure 2B, D. Curves on the left show training and testing on matched timepoints. Matrices on the right show generalization of patterns across all pairs of training and testing time windows.

Across the whole sensor space (Figure 2B), significant discrimination between the two auditory stimuli/cues emerged shortly after the presentation of the stimulus, peaking at around 108 ms for the auditory localizer task, 116 ms for the preparatory phase of the attention task, and 108 ms when training the classifier on the localizer task and testing on the attention task. During the attention task, cue discrimination then returned to chance level. Decoding during the preparatory phase of the attention task was initially largely driven by low-level auditory differences between cues, shown by cross-generalization of patterns between the attention task and the auditory localizer task (Figure 2B, yellow curve). However, the decoding peak during the attention task (Figure 2B, blue curve) was higher than in the auditory localizer task (Figure 2B, red curve), and higher than the decoding that could be accounted for by patterns shared between the tasks (Figure 2B, yellow curve).

To explore this further, we compared the decoding time-courses in source space (Figure 2D). In the visual cortex and LPFC, but not in the auditory cortex, stronger decoding was observed

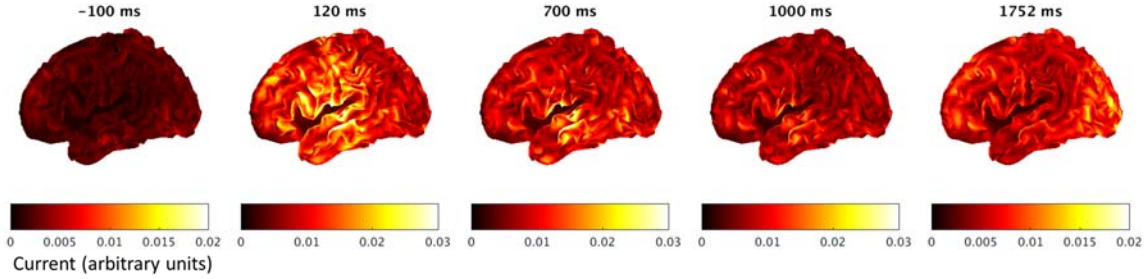
during the preparatory phase of the attention task than during the auditory localizer task. Matched decoding accuracy for all three curves in auditory cortex suggests that the representation in this region is purely sensory. Stronger and broader decoding peaks for the attention condition in visual cortex and LPFC suggests that these regions are representing information additional to low-level auditory processing. This increased representation may reflect the additional representation of contextual condition of the attention task, required to associate the auditory stimulus with its visual target. These data also imply that decoding in frontal/visual regions is not simply a result of signal leakage from auditory cortex.

Towards the end of the preparatory phase of the attention task, just before the visual stimuli appeared, weak decoding of the cue reoccurred in the LPFC, which had also been seen in sensor space. Notably, this decoding appeared immediately before the first possible time (1750 ms) at which the first choice stimulus might appear. Although this signal might reflect template reactivation immediately before a choice was anticipated (Myers et al., 2015; Kok et al. 2017; Grubert and Eimer, 2018), it was extremely weak.

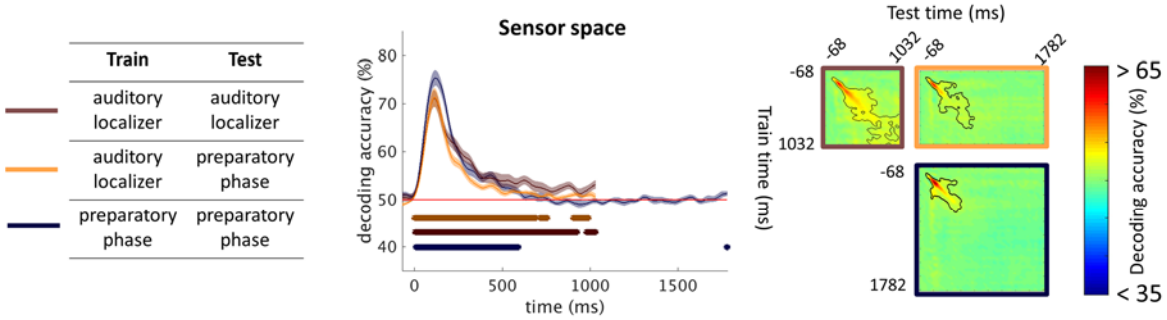
To test whether activity during the preparatory phase might reflect the representation of the upcoming trial target, we performed a cross-task and cross-time classification analysis using the visual localizer task. At every time window, patterns from the two visual items associated with each cue were taken from the visual localizer task to use as training data, and these were tested at every time window of the preparatory phase of the attention task to decode the trial target. We did not find any significant clusters where the visual template cross-generalized to the preparatory phase.

Finally, we note that cross-time generalization matrices suggest that the LPFC signal reached a steady state at the end of the auditory localizer, in contrast to its lack of sustained signal during the preparatory phase of the attention task. This might reflect that the fact that the representation in the auditory task does not need to be transformed further, whereas in the attention task it serves an intermediate role in mapping subsequent visual inputs to behavior.

A. Source localization of attentional cue



B. Sensor space decoding



C. Regions of interest

D. Source space decoding

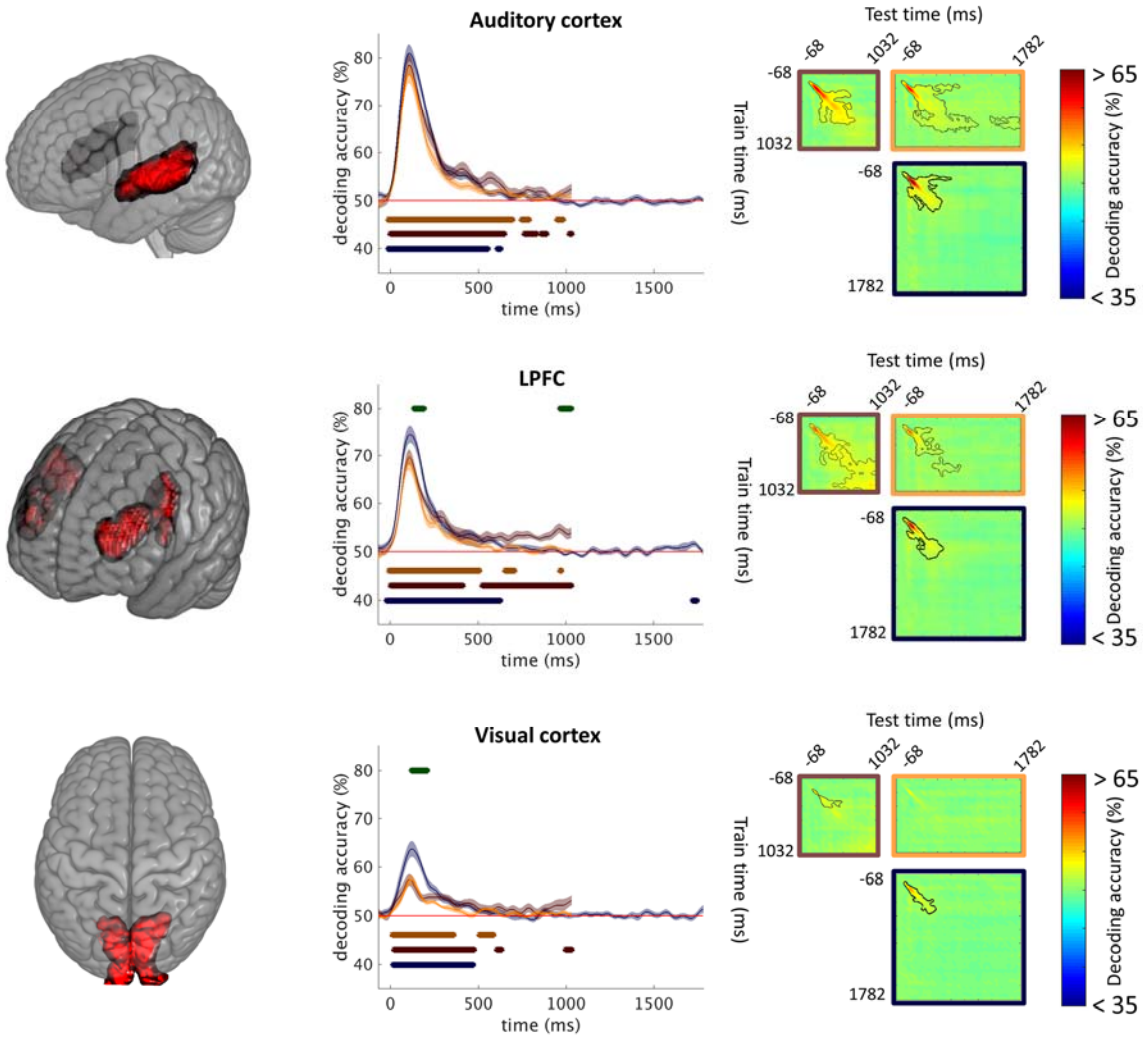


Figure 2. Response to the attentional cue. (A) Source localization of EEG/MEG response to the auditory attentional cue at representative time points relative to cue onset. (B) Decoding time-course of auditory stimulus/attentional cue using all sensors combining EEG and MEG across the whole brain. Curves on the left show decoding when training and testing on matched time-points. Colored dots beneath the decoding curves show times where decoding is significantly above chance for each condition. Green dots above indicate times where classification accuracy within the attention task is significantly different from the mean of the auditory-task and cross- task classification accuracies. Significance is corrected for multiple comparisons across time (FDR < 0.05). Translucent bands represent standard error of the mean. Matrices on the right show temporal generalization of decoding across all pairs of training and testing times. Black contours indicate regions of significant decoding (p<0.05) after cluster correction. (C) Vertices within source space ROIs (auditory cortex, lateral prefrontal cortex (LPFC), and visual cortex). (D) Decoding time-courses from these source space ROIs; same format as (B).

Coding of visual and behavioral properties of 1-item displays

We next turned to processing of the visual items, and selection of the target item. Source localization of the response to the visual stimuli at representative time points is shown in figure 3A.

During 1-item displays, we expected strong, early discrimination of object identity (e.g., face vs house). In the attention task, each stimulus additionally had a behavioral category depending on the cue of that trial. For the participant to make the appropriate response to each stimulus, we expected that the neural signal would also show behavioral category discrimination (target vs. non-target), which would occur after object identity processing. For these analyses, we focused on the T and Ni conditions, for which object identity and behavioral category were fully crossed. Object representation was measured by the

discrimination between faces and houses when each were equally often targets or non-targets; behavioral category representation was measured by discrimination between targets and non-targets when each were equally often faces or houses.

Single stimulus decoding time-courses on T and Ni presentations are shown in Figure 3B-C. In line with expectations, both object identity and behavioral category showed substantial periods of significant decoding accuracy. Across the whole sensor space, a significant difference between object identities peaked at around 132 ms. Behavioral category decoding emerged later, slowly rising to a peak at 344 ms.

Source space analysis showed that both types of information could be decoded from all three ROIs. Decoding of object identity in the auditory ROI warns of possible signal leakage between regions. Visual cortex, however, had the highest decoding accuracy for object identity, while ROIs did not statistically differ in their strength of decoding accuracy for behavioral category.

Cross-temporal generalization indicated that object identity representation was most stable in the visual ROI, but decayed to baseline in all ROIs. In contrast, behavioral category representation appeared to settle into a steady state in the LFPC ROI.

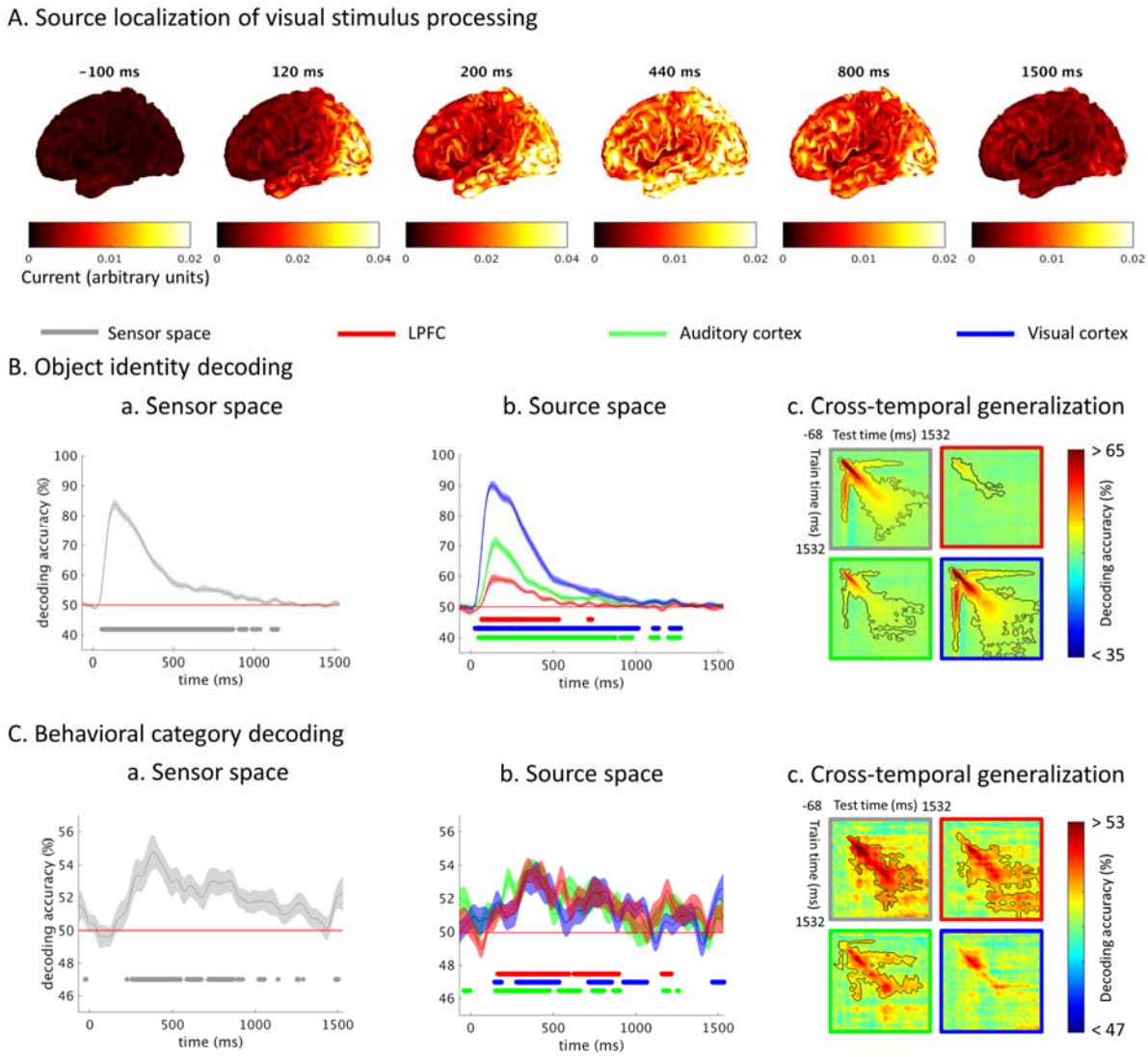


Figure 3. Coding of visual and behavioral properties of 1-item displays (A) Source localization of EEG/MEG response to visual presentation (including both single-item and multi-item displays) at representative time-points. (B) Decoding time-courses of object identity, in (a) sensor and (b) source space, when training/testing using matched time-points, and (c) generalizing across training/testing times. Dots beneath the decoding curves show times where decoding is significantly above chance, corrected for multiple comparisons across time (FDR<0.05). Translucent bands represent standard error of the mean. Black contours in cross-time matrices indicate regions of significant decoding (p<0.05) after cluster correction. (C) Decoding time-courses and cross-temporal generalization for behavioral category information. Object identity decoding emerged earlier than behavioral category decoding. Visual cortex showed the highest object decoding accuracy, while ROIs were comparable in their strength of behavioral category representation.

Coding of target location in 3-item displays

Next, we examined target representation in the presence of simultaneous distractors. We first asked whether the spatial location of the target within 3-item displays could be decoded (Figure 4). To do this, we decoded every pair of T versus Ni locations, while holding Nc position constant (i.e., “T right, Ni left” vs. “Ni right, T left”, “T right, Ni bottom” vs. “Ni right, T bottom”, and “T left, Ni bottom” vs. “Ni left, T bottom”) and averaged the accuracies within each participant. Group sensor-space results showed that decoding began to emerge shortly after stimulus onset at 72 ms, and peaked at 288 ms, before slowly declining toward the end of the epoch. The analysis was repeated in source space. Decoding of target location was significant in all ROIs, but strongest in visual cortex. Cross-temporal generalization suggested that the representation of target location decayed rather than settling into a steady state.

In a complementary analysis to target location decoding, we examined the N2pc, a well-known early index of spatial attention, which appears as a negativity over posterior EEG electrodes contralateral to the side of space to which the subject is attending around 200-300 ms following a stimulus (Luck and Hillyard, 1994; Heinze et al., 1990; Eimer, 1996; Hopf et al. 2000). We compared event-related potentials/fields when the target was on the right or left of the screen of the 3-item display, and the topography of this contrast is shown in Figure 4B. Differences between target locations peaked between 200-300 ms in posterior EEG and MEG signals.

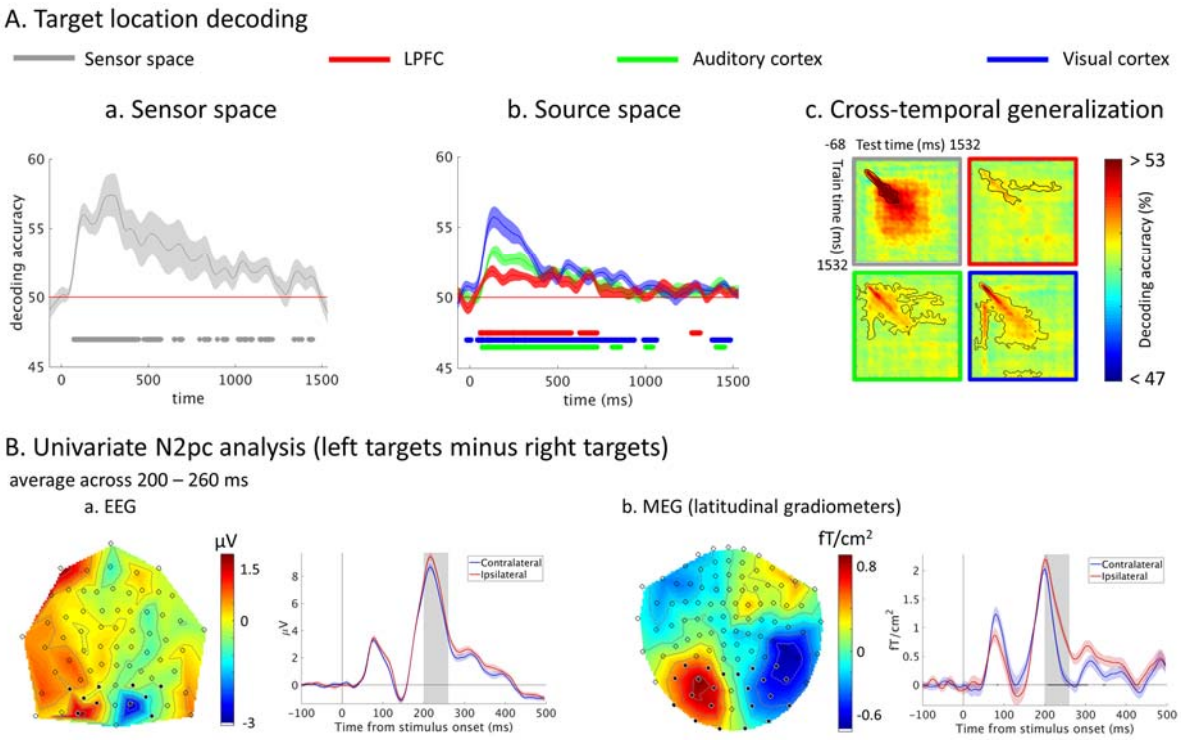


Figure 4. Coding of target location in 3-item displays (A) Decoding of target location during presentation of 3-item displays, i.e., whether the item corresponding to the cue is in the left, right, or bottom position. Format as in Figure 3. Location decoding was strongest in the visual cortex. (B) Univariate N2pc ERP/ERFs across (a) EEG electrodes and (b) latitudinal gradiometers. Latitudinal gradiometers are presented because their orientation around the helmet means that contralateral asymmetries in the magnetic flux gradient are expressed analogously to the EEG topography (Mitchell and Cusack, 2011; Kuo et al. 2016). Topographies are averaged across 200-260 ms (marked in grey on the timecourses). Timecourses are averaged across posterior sensors contralateral and ipsilateral to the target (highlighted on the topographies), with black dots indicating a significant difference (p<0.05) after FDR correction.

Coding of target identity during presentation of 3-item displays

We also hypothesized that representation of 3-item displays would differ depending on the cue, even though the visual input is the same. All 3-item displays contained the target item that was associated with the cue as well as the Ni and Nc items. Therefore, the decoding of

the cue in the presence of a matching visual stimulus likely reflects attentional enhancement of the selected target identity. In sensor space, cue/target identity decoding emerged from around 212 ms and peaked at 252 ms. In source space, the visual cortex showed the highest decoding accuracy (Figure 5).

Similar to the representation of location, cross-temporal generalization suggested that the representation of target identity in the presence of distractors decayed rather than settling into a steady state. For this analysis, we also expected cross-task generalization from the visual localizer. This was significant in the visual ROI, but not in the auditory or LPFC ROIs, suggesting that the shared pattern was predominantly sensory, with minimal signal leakage in this case.

To compare the decoding latencies of target location and target identity in 3-item displays, the time of peak sensor space as well as visual cortex decoding of each type was identified for each subject. If either peak fell outside of a 0 – 600 ms window for a given subject, then that subject was removed from the analysis. Paired t-tests showed that the timing of peak target location decoding and target identity decoding did not differ significantly in the whole sensor space ($t(10) = 1.18$, $p = 0.26$). However, target location decoding preceded target identity decoding ($t(14) = 2.99$, $p < 0.01$) in the visual cortex.

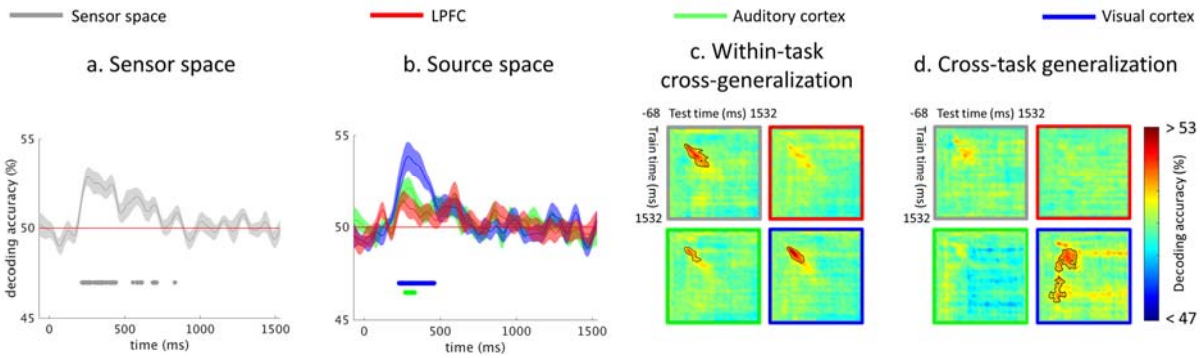


Figure 5. Decoding of attentional cue/target identity during presentation of 3-item displays. Panels (a-c) have the same format as Figure 3. Panel (d) shows cross-task generalization of decoding, when training on the visual localizer task and testing on the attention task.

Reawakening of the attentional cue/template during presentation of consistent non-targets

Finally, we tested whether we could decode the cue/template during the presentation of a single Nc visual stimulus. Wolff et al. (2015, 2017) have shown that by 'pinging' the brain with a neutral stimulus during working memory maintenance, it can be possible to decode the memory-item-specific information from the impulse response. In our data, cue decoding during Nc presentation was visible but rather weak (Figure 6). Across sensor space and source space, there were scattered brief periods of significant decoding. Their appearance in auditory as well as visual and frontal ROIs questions whether these might reflect a reactivated memory of the auditory cue, or a visual attentional template in anticipation of the next visual input. Apparent signal in the auditory ROI might also reflect leakage from other sources. Like the analysis of activity in the post-cue delay, these data indicate that representations consistent with an attentional template were weak and intermittent. Cross-temporal generalization suggested that although the representation was not fully sustained, when it resurfaced in the visual ROI it did so with a similar pattern. Cross-task generalization from the auditory and visual localizers provided no evidence that this representation was in a similar format to either cue or target perception.

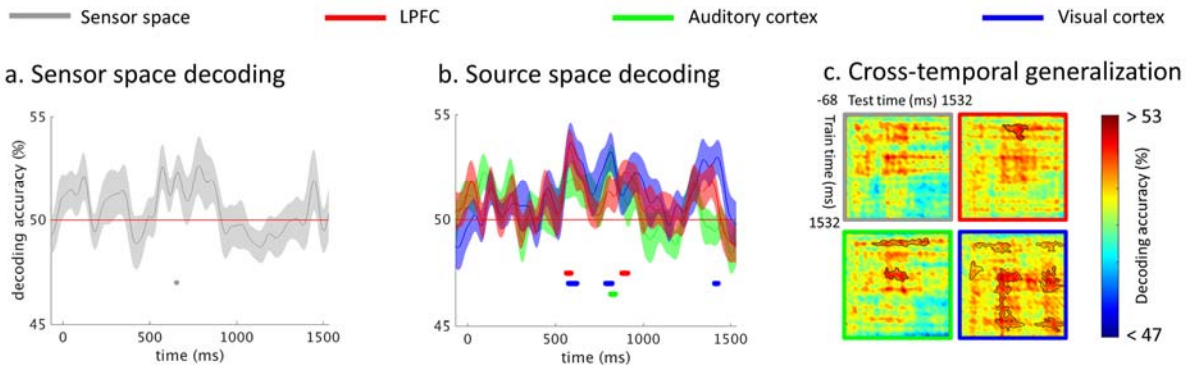


Figure 6. Decoding timecourse of attention cue during presentation of Nc displays. Format as in Figure 3.

Summary of component time-courses during attentional selection

Above we have described five distinct forms of information representation evoked by the appearance of the visual stimuli (Figures 3-6). These are summarized in Figure 7, overlaying their average sensor-space and ROI-based decoding time-courses for ease of comparison.

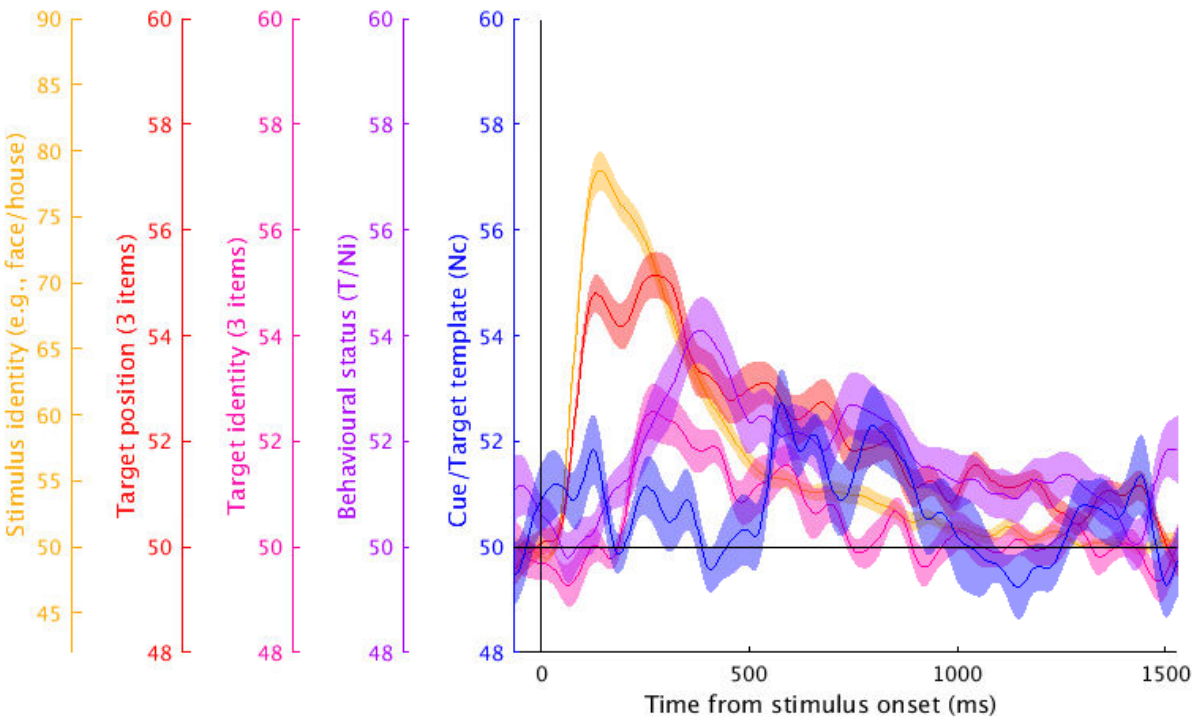


Figure 7. Summary of the decoding time-courses of five component processes of selective attention following onset of a visual stimulus: representation of stimulus identity, target position, target identity, behavioral status, and the template of the cue/target.

Discussion

There is currently much interest in decoding the contents of cognitive operations from human MEG/EEG data. Here, we examined the evolution of multiple forms of information represented in the brain as a visual target is selected. Combining single-item displays with multi-item displays of targets and different types of distractors allowed quantification of distinct components of processing during selective attention, indexed by different profiles of representational content.

Selection of a particular target must begin from an attentional template, specifying the object relevant to current behavior (Duncan and Humphreys, 1989; Bundesen, 1990). In fMRI, multivariate classifiers trained on responses to viewed stimuli can predict an attentional template during the preparatory phase (e.g., Stokes et al., 2009). In our MEG/EEG data, both visual and frontal ROIs showed greater decoding of cue identity in the attention task compared to the auditory localizer. Plausibly, this reflects processes of establishing an attentional template or rule for subsequent behavior. However, beyond 1000 ms, cue decoding was indistinguishable from chance, except for a minimal hint of reappearance, in sensor space and the frontal ROI, immediately before the anticipated stimulus onset. Previous EEG/MEG studies have suggested existence of a pre-stimulus template, also often subtle and short-lived (Myers et al., 2015; Kok et al., 2017; Grubert and Eimer, 2018). Following non-

target displays, we observed evidence of template reawakening; again, this was weak and not fully sustained.

Sustained activity reflecting an attentional template may be largely invisible to MEG/EEG, perhaps because it is dynamic, with variability in timing or format across trials. Relatively weak template coding is consistent with direct neural recordings in behaving monkeys (Chelazzi et al., 1993; Bichot et al., 2015). A parallel possibility is that attentional templates may sometimes be stored in a passive form, such as changed synaptic weights (Stokes, 2015). Substantial variability in template representation is directly predicted by integrated competition accounts of attention (Duncan et al. 1997), which propose that an entire object comes to dominate the mental representation regardless of which features provide the initial attentional bias. With the complex real-world objects used here, any one of numerous features could be sufficient to bias processing toward selection of the target. Since the particular target feature can vary freely across trials and individuals, it is perhaps unsurprising that we see only weak evidence of consistent template representation, and no evidence that its format matches full target perception. Indeed, even when an object is imagined in detail, only a subset of its visually-evoked feature are instantiated, and these can vary considerably across people (Mitchell and Cusack, 2016). We predict that the extent to which perceptually veridical template representations are seen will reflect constraints on which features can effectively bias integrated competition. Consistent templates are more likely when targets are simple or few features distinguish targets from distractors (Kok et al; Myers et al; Grubert and Eimer, 2018). Perceptually complete templates are more likely when targets share different features with different distractors (Duncan and Humphreys, 1989). Anticipation of stimulus timing may also influence template representation, with templates activated/strengthened only when the search display is imminent (Grubert and Eimer 2018).

Upon presentation of the visual choice display, we found much decodable information of various kinds. The timing of peak decoding of different features suggests five components of processing. The current data cannot determine the extent to which these components evolve in parallel or have some serial dependency, whereby one process influences another. It is likely that there is a degree of both (Bichot et al., 2005). First, visual stimulus properties are encoded, shown by object identity decoding in 1-item displays, peaking around 132 ms, and strongest in visual cortex. Second, in a multielement display, the candidate target is localized, shown by target location decoding that peaked between 136 ms (in visual cortex, where strongest) and 288 ms (combining all sensors). This may be partially concurrent with initial visual processing, consistent with an initial parallel stage of selection (Duncan, 1980; Treisman and Gelade, 1980) and a special role for space in feature binding (Itti and Koch, 2000; Bisley and Goldberg, 2010); Schneegans and Bays, 2017). Third, representation of the candidate target continues to be enhanced relative to distractors, perhaps via integrated competition, shown by cue/target identity decoding in 3-item displays, peaking around 252 ms, again strongest in visual cortex. Fourth, behavioral significance of the target is explicitly represented (in this case whether it is a target, so requiring further processing), shown by behavioral category decoding in 1-item displays, peaking around 344 ms and most stable in the LPFC. Fifth, if no target is identified and search must continue, an attentional template might be reactivated or strengthened, shown by cue decoding after Nc displays, peaking beyond 500 ms and appearing generally widespread. The precise timing at which each representation is detectable will depend on many factors including stimuli, task, analysis sensitivity, similarity between targets and distractors, and the number and homogeneity of distractors (Duncan and Humphreys 1989). Nonetheless, we anticipate that the sequence of key components would largely generalize across paradigms. Potential dependencies between

processes might be investigated by combining MVPA of electrophysiological recordings with transcranial magnetic stimulation at successive times.

In 1-item displays, we found a distinction between visual cortex and LPFC. While the regions represented behavioral category with similar strength, visual cortex represented stimulus identity more strongly than LPFC. Similarly, object identity was represented more stably in visual cortex, whereas behavioral category was represented more stably in LPFC. fMRI studies show that frontal regions flexibly code for behaviorally relevant categories according to task rule (Jiang et al., 2007; Li et al., 2007; Woolgar et al., 2011; Erez and Duncan, 2015). Electrophysiological recordings of monkey prefrontal responses to T, Ni, and Nc stimuli show that visual input properties are initially equally represented for targets and non-targets, whereas the behaviorally critical target dominates later processing (Kadohisa et al., 2013; Stokes et al., 2013). Our results also suggest an anterior-posterior distinction in information content and timing.

Interestingly, a target influenced bias in the 3-item displays well before its behavioral relevance was explicitly decodable in the single-item displays. This strongly suggests at least two stages of selection. Distinction between an early parallel stage and a later capacity-limited stage is central to most models of attention (Duncan, 1980; Treisman & Gelade, 1980). Target decoding in 3-item displays emerged from 212 ms, consistent with the earliest attentional modulation of stimulus category processing in cluttered scenes observed from 180 ms (Kaiser et al., 2016), and with prior demonstration of feature-binding during integrated competition at a similar timescale (Schoenfeld et al. 2003). The later stage may correspond to capacity-limited individuation of the integrated target object, allowing its bound properties to become accessible for further processing and goal-directed action (Duncan, 1980; Bichot et

al., 2005; Mitchell and Cusack 2007 Christie et al., 2015), in this case likely including the brightness judgement. These two stages could also be interpreted in terms of the "global neuronal workspace" model - the earlier attentional bias reflecting accumulation of pre-conscious sensory evidence; the later explicit representation of target status reflecting conscious awareness and "ignition" of fronto-parietal networks, linked to P3 waves around 300- 500 ms (Dehaene & Changuex, 2011; Sergent et al., 2005). Although target localization implies target identification, and time-courses of location and identity representation in 3-item displays were heavily overlapping, the location signal peaked first. This might arise because feature integration is strongly mediated by shared location within spatiotopic maps (Schneegans and Bays, 2017), and indicates that the early phase of integrated competition is dynamic, enhancing different aspects of the target representation at different times. Furthermore, neither of these competitive representations of target features appeared to reach a steady state. In contrast, explicit representation of target status settled into a steady state, as did object identity in the absence of simultaneous distractors.

To conclude, although attentional selection must begin with a template, this may be weakly or variably represented (Duncan et al. 1997), or even maintained in "silent" form (Stokes, 2015). In agreement with others (Olivers et al., 2011; Myers et al., 2015; Grubert and Eimer, 2018), we suggest that the template may be actively represented only when needed, and least likely to interfere with other concurrent processes. Integrated competition accounts of attention imply that the template need be neither complete nor constant across trials, consistent with no significant response pattern generalization between template representations and the visual localizer. In contrast, integrated competition suggests that attentional selection and enhancement of stimulus representations will be strong and widespread. Supporting such models, we observed robust decoding of the critical processing

stages required to select and enhance a target amongst competing distractors, and to categorize it according to behavioral requirements.

References

Beck DM, Kastner S (2009) Top-down and bottom-up mechanisms in biasing competition in the human brain. Vision Res 49:1154–1165.

Bichot NP, Heard MT, DeGennaro EM, Desimone R (2015) A Source for Feature-Based Attention in the Prefrontal Cortex. Neuron 88:832–844.

Bichot NP, Rossi AF, Desimone R, Bichot NP, Rossi AF, Desimone R (2005) Parallel and Serial Neural Mechanisms for Visual Search in Macaque Area V4. Science (80-) 308:529–534.

Bisley JW, Goldberg ME (2010) Attention, Intention, and Priority in the Parietal Lobe. Annu Rev Neurosci 33:1–21.

Brainard DH, Vision S (1997) The psychophysics toolbox. Spatial vision, 10, 433-436.

Bundesen C (1990) A theory of visual attention. Psychol Rev 97:523–547.

Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM transactions on intelligent systems and technology (TIST), 2(3), 27.

Chelazzi L, Miller EK, Duncan J, Desimone R (1993) A neural basis for visual search in inferior temporal cortex. Nature 363:345–347.

Christie GJ, Livingstone AC, McDonald JJ (2015) Searching for Inefficiency in Visual Search. J Cogn Neurosci 27:46–56.

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(1), 9-21.

Dehaene S, Changeux JP (2011) Experimental and Theoretical Approaches to Conscious Processing. Neuron 70:200–227.

Desimone R, Duncan J (1995) Neural Mechanisms of Selective Visual. Annu Rev Neurosci 18:193–222.

Duncan J (1980) The locus of interference in the perception of simultaneous stimuli. Psychol Rev 87:272–300.

Duncan J, Humphreys G, Ward R (1997) Competitive brain activity in visual attention. Curr Opin Neurobiol 7:255–261.

Duncan J, Humphreys GW (1989) Visual search and stimulus similarity. Psychol Rev 96:433–458.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage, 25(4), 1325-1335.

Eimer M (1996) The N2pc component as an indicator of attentional selectivity. Electroencephalography and clinical neurophysiology, 99(3), 225-234.

Erez Y, Duncan J (2015) Discrimination of Visual Categories Based on Behavioral Relevance in Widespread Regions of Frontoparietal Cortex. J Neurosci 35:12383–12393.

Fedorenko E, Duncan J, Kanwisher N (2013) Broad domain generality in focal regions of frontal and parietal cortex. Proceedings of the National Academy of Sciences, 201315235.

Grootswagers T, Wardle SG, Carlson TA (2017) Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. Journal of cognitive neuroscience, 29(4), 677-697.

Grubert A, Eimer M (2018) The time course of target template activation processes during preparation for visual search. Journal of Neuroscience, 38(44), 9527-9538.

Hebart MN, Bankson BB, Harel A, Baker CI, Cichy RM (2018) The representational dynamics of task and object processing in humans. Elife 7.

Heinze HJ, Luck SJ, Mangun GR, Hillyard SA (1990) Visual event-related potentials index focused attention within bilateral stimulus arrays. I. Evidence for early selection. Electroencephalography and clinical neurophysiology, 75(6), 511-527.

Hopf JM, Luck SJ, Girelli M, Hagner T, Mangun GR, Scheich H, Heinze HJ (2000) Neural sources of focused attention in visual search. Cereb Cortex 10:1233–1241.

Itti L, Koch C (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. In: Vision Research, pp 1489–1506.

Jiang X, Bradley E, Rini RA, Zeffiro T, VanMeter J, Riesenhuber M (2007) Categorization Training Results in Shape- and Category-Selective Human Neural Plasticity. Neuron 53:891–903.

Kaiser D, Oosterhof NN, Peelen MV (2016) The neural dynamics of attentional selection in natural scenes. Journal of neuroscience, 36(41), 10522-10528.

Kadohisa M, Petrov P, Stokes M, Sigala N, Buckley M, Gaffan D, Kusunoki M, Duncan J (2013) Dynamic Construction of a Coherent Attentional State in a Prefrontal Cell Population. Neuron 80:235–246.

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

Kok P, Brouwer GJ, van Gerven MAJ, de Lange FP (2013) Prior Expectations Bias Sensory Representations in Visual Cortex. J Neurosci 33:16275–16284.

Kok P, Mostert P, de Lange FP (2017) Prior expectations induce prestimulus sensory templates. Proc Natl Acad Sci:201705652.

Kuo BC, Nobre AC, Scerif G, Astle DE (2016) Top–Down Activation of Spatiotopic Sensory Codes in Perceptual and Working Memory Search. Journal of cognitive neuroscience, 28(7), 996-1009.

Li S, Ostwald D, Giese M, Kourtzi Z (2007) Flexible Coding for Categorical Decisions in the Human Brain. J Neurosci 27:12321–12330.

Luck SJ, Hillyard SA (1994) Spatial filtering during visual search: evidence from human electrophysiology. J Exp Psychol Hum Percept Perform 20:1000–1014.

Mitchell DJ, Cusack R (2007) Flexible, capacity-limited activity of posterior parietal cortex in perceptual as well as visual short-term memory tasks. Cerebral Cortex, 18(8), 1788-1798.

Mitchell DJ, Cusack R (2011) The temporal evolution of electromagnetic markers sensitive to the capacity limits of visual short-term memory. Frontiers in human neuroscience, 5, 18.

Mitchell DJ, Cusack R (2016) Semantic and emotional content of imagined representations in human occipitotemporal cortex. Scientific reports, 6, 20232.

Myers NE, Rohenkohl G, Wyart V, Woolrich MW, Nobre AC, Stokes MG (2015) Testing sensory evidence against mnemonic templates. Elife 4:1–25.

Olivers CN, Peters J, Houtkamp R, Roelfsema PR (2011) Different states in visual working memory: When it guides attention and when it does not. Trends in cognitive sciences, 15(7), 327-334.

Peelen M V., Kastner S (2011) A neural basis for real-world visual search in human occipitotemporal cortex. Proc Natl Acad Sci 108:12125–12130.

Puri AM, Wojciulik E, Ranganath C (2009) Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. Brain Res 1301:89–99.

Schoenfeld MA, Tempelmann C, Martinez A, Hopf JM, Sattler C, Heinze HJ, Hillyard SA

(2003) Dynamics of feature binding during object-selective attention. Proceedings of the National Academy of Sciences, 100(20), 11806-11811.

Schneegans S, Bays PM (2017) Neural architecture for feature binding in visual working memory. Journal of Neuroscience, 3493-16.

Sergent C, Baillet S, Dehaene S (2005) Timing of the brain events underlying access to consciousness during the attentional blink. Nat Neurosci 8:1391–1400.

Stokes M, Thompson R, Nobre AC, Duncan J (2009) Shape-specific preparatory activity mediates attention to targets in human visual cortex. Proc Natl Acad Sci 106:19569–19574.

Stokes MG (2015) "Activity-silent" working memory in prefrontal cortex: A dynamic coding framework. Trends Cogn Sci 19:394–405.

Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J (2013) Dynamic coding for cognitive control in prefrontal cortex. Neuron 78:364–375.

Stokes MG, Wolff MJ, Spaak E (2015) Decoding Rich Spatial Information with High Temporal Resolution. Trends Cogn Sci 19:636–638.

Taulu S, Simola J (2006) Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Physics in Medicine & Biology, 51(7), 1759.

Theeuwes J (2010) Top-down and bottom-up control of visual selection. Acta Psychol (Amst) 135:77–99.

Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cognitive psychology, 12(1), 97-136.

Wolff MJ, Ding J, Myers NE, Stokes MG (2015) Revealing hidden states in visual working memory using electroencephalography. Front Syst Neurosci 9:1–12.

Wolff MJ, Jochim J, Akyürek EG, Stokes MG (2017) Dynamic hidden states underlying

working-memory-guided behavior. Nat Neurosci, *20*(6), 864.

Woolgar A, Hampshire A, Thompson R, Duncan J (2011) Adaptive Coding of Task-Relevant Information in Human Frontoparietal Cortex. J Neurosci 31:14592–14599.

Wutz A, Loonis R, Roy JE, Donoghue JA, Miller EK (2018) Different Levels of Category Abstraction by Different Dynamics in Different Prefrontal Areas. Neuron, *97*(3), 716-726.