High temporal resolution decoding of object position and category

Thomas A. Carlson

Department of Psychology, University of Maryland, College Park, MD, USA

Hinze Hogendoorn

Helmholtz Institute, Psychonomics Division, Universiteit Utrecht, The Netherlands

Ryota Kanai

Institute of Cognitive Neuroscience and Department of Psychology, University College London, London, UK

Juraj Mesik

Department of Psychology, University of Maryland, College Park, MD, USA

Jeremy Turret

Department of Psychology, University of Maryland, College Park, MD, USA

We effortlessly and seemingly instantaneously recognize thousands of objects, although we rarely—if ever—see the same image of an object twice. The retinal image of an object can vary by context, size, viewpoint, illumination, and location. The present study examined how the visual system abstracts object category across variations in retinal location. In three experiments, participants viewed images of objects presented to different retinal locations while brain activity was recorded using magnetoencephalography (MEG). A pattern classifier was trained to recover the stimulus position (Experiments 1, 2, and 3) and category (Experiment 3) from the recordings. Using this decoding approach, we show that an object's location in the visual field can be recovered in high temporal resolution (5 ms) and with sufficient fidelity to capture topographic organization in visual areas. Experiment 3 showed that an object's category could be recovered from the recordings as early as 135 ms after the onset of the stimulus and that category decoding generalized across retinal location (i.e., position invariance). Our experiments thus show that the visual system rapidly constructs a category representation for objects that is invariant to position.

Introduction

The same object can generate an infinite number of retinal images through variations in size, pose, illumination, and location. Still, we effortlessly recognize hundreds of objects under these highly variable conditions. Successful recognition necessitates a higher order representation that is invariant, or sufficiently tolerant, to these variations. Position invariance is the capacity to abstract category and exemplar information across changes in retinal position. It is well known that the retinal location of a stimulus is initially encoded in topographic maps in early visual areas (Engel, Glover, & Wandell, 1997; Sereno et al., 1995; Wandell, Dumoulin, & Brewer, 2007). Recent human imaging studies have shown that ventral visual areas, which are often described as being object-selective

(Malach et al., 1995) or category-selective (Downing, Jiang, Shuman, & Kanwisher, 2001; Epstein & Kanwisher, 1998; Kanwisher, McDermott, & Chun, 1997), also represent an object's position in the visual field (Grill-Spector et al., 1999; Hemond, Kanwisher, & Op de Beeck, 2007; MacEvoy & Epstein, 2007; Niemeier, Goltz, Kuchinad, Tweed, & Vilis, 2005).

Our capacity to recognize objects is mediated by a constellation of areas in the ventral visual pathway (Ungerleider & Mishkin, 1982). Theories of object recognition need to account for position coding in ventral temporal cortex (Kravitz, Vinson, & Baker, 2008), as they would also be subject to the same limitation as retinal representation—that is, an object seen at different locations in the visual field will elicit different patterns of activation. DiCarlo and Cox (2007) theorize that invariance can be achieved by ensembles of neurons encoding multiple stimulus properties

in a manner allowing for category and identity to be read out. Contemporary research has supported this theory by using multivariate methods to study patterns of activation in high dimensional space. Hung, Kreiman, Poggio, and DiCarlo (2005), for example, used pattern classification techniques to decode invariant identity and category information from populations of neurons in primate inferior temporal cortex. Subsequent fMRI studies have provided converging evidence demonstrating that exemplar and category information can be decoded from ventral visual areas that also represent the position of a stimulus (Carlson, Hogendoorn, Fonteijn, & Verstraten, 2011; Cichy, Chen, & Haynes, 2011; Sayres & Grill-Spector, 2008; Schwarzlose, Swisher, Dang, & Kanwisher, 2008; but see Kravitz, Kriegeskorte, & Baker, 2010).

The time that at which the brain constructs an invariant representation of an object is a useful constraint on models of object recognition that invoke feed-forward (Fukushima, 1980; Riesenhuber & Poggio, 1999; Rolls, 1991; Serre, Oliva, & Poggio, 2007; VanRullen & Thorpe, 2002) and recurrent mechanisms (Bullier, 2001; Garrido et al., 2007; Hochstein & Ahissar, 2002). Generally speaking, rapid construction of invariant object representations favors feed-forward models, as there is arguably insufficient time for recurrent processes. This is illustrated by two recent studies examining the temporal dynamics of invariance. Liu et al. (2009), using human intracranial recordings, found that invariant object representations of scale and depth could be decoded as early as 100 ms post-stimulus onset. The authors accordingly interpreted their findings as support for feed-forward models due to the short latency. In contrast, Freiwald & Tsao (2010) found that invariance to viewpoint arose much later (300 ms) and concluded that computations of view invariance involved recurrent processing.

The present study examined the temporal dynamics of position coding and translation invariance using MEG. We used pattern classification analysis to decode the position of an object in the visual field and its category from the recordings. While this decoding approach has been used extensively in physiology and fMRI, it has been relatively underutilized in EEG and MEG. In Experiments 1 and 2, we demonstrate the utility of this approach by showing that object position can be recovered in high temporal resolution (5 ms) from single trial data. Moreover, classification performance was well described by known topographic organization in visual areas, thus demonstrating the sensitivity of the approach. In our third experiment, we used this decoding approach to show that location-invariant responses to objects arise as early as 110 ms post-stimulus onset. As a final aside, we observed a strong neurophysiological response to the offset of the stimulus that similarly encodes the position and category of the stimulus that was no longer physically present. Given the robustness and informative nature of the offset response, we hypothesize a functional role for visual offsets in short-term memory.

Methods and materials

Participants

Eight volunteers with normal or corrected-to-normal vision participated in three experiments (mean age 23, range 19–36). Subjects were paid for their participation. Experimental procedures were approved by the University of Maryland Institutional Review Board.

Display apparatus

Subjects were supine in the MEG recording chamber. Stimuli were displayed on a translucent screen, located approximately 50 cm above the subject, by a projector located outside of the magnetically shielded recording chamber. Stimuli were displayed using a Dell PC desktop computer running Presentation software (Neurobehavioral Systems; http://www.neurobs.com).

Stimuli

In Experiments 1 and 2, stimuli were images of faces (8 males, 8 females) taken from the Psychological Image Collection at Stirling (http://pics.psych.stir.ac.uk). Experiment 3 had four categories of images including faces, cars, and two artificial texture categories. The images of cars (16 total) were cropped photographs of automobiles taken at the University of Maryland. Artificial textures were generated using a texture synthesis procedure that replicates local image statistics (Portilla & Simoncelli, 2000; for an extended discussion of the advantages of this procedure, see Rust & Dicarlo, 2010). The synthesis procedure was run on each face and car image to produce two additional categories: face textures (16 total) and car textures (16 total). The texture images were used in Experiment 3 to control for differences in low-level properties, e.g., contrast and spatial frequency, between the face and car images.

Stimuli were displayed on a uniform gray background. The images were cropped and displayed in circular aperture that measured 9 degrees of visual angle in diameter (Figure 1A). Stimuli were presented 7 degrees of visual angle into the periphery. Images were presented in the lower visual field to maximize the evoked response (see Fylan, Holliday, Singh, Anderson, & Harding, 1997; Poghosyan & Ioannides, 2007; Portin, Vanni, Virsu, & Hari, 1999). The position of the images was varied by changing the angular offset of the image relative to a central position on the lower vertical meridian (Figure 1B). Experiments 1 and 2 tested seven positions with offsets of −60, −30, −15, 0, 15, 30, and 60 degrees. In Experiment 3, the number of positions was reduced to three to accommodate

the additional number of trials necessary to test the four categories of stimuli. Stimuli in Experiment 3 were presented at offsets of −45, 0, and 22.5 degrees. This irregular spacing was used to increase the number of unique pairwise comparisons for the position analysis.

Experimental design

Figure 1C shows a diagram of a typical trial. Each trial consisted of the following sequence of events: A fixation cross was displayed for 300 ms; the first image was
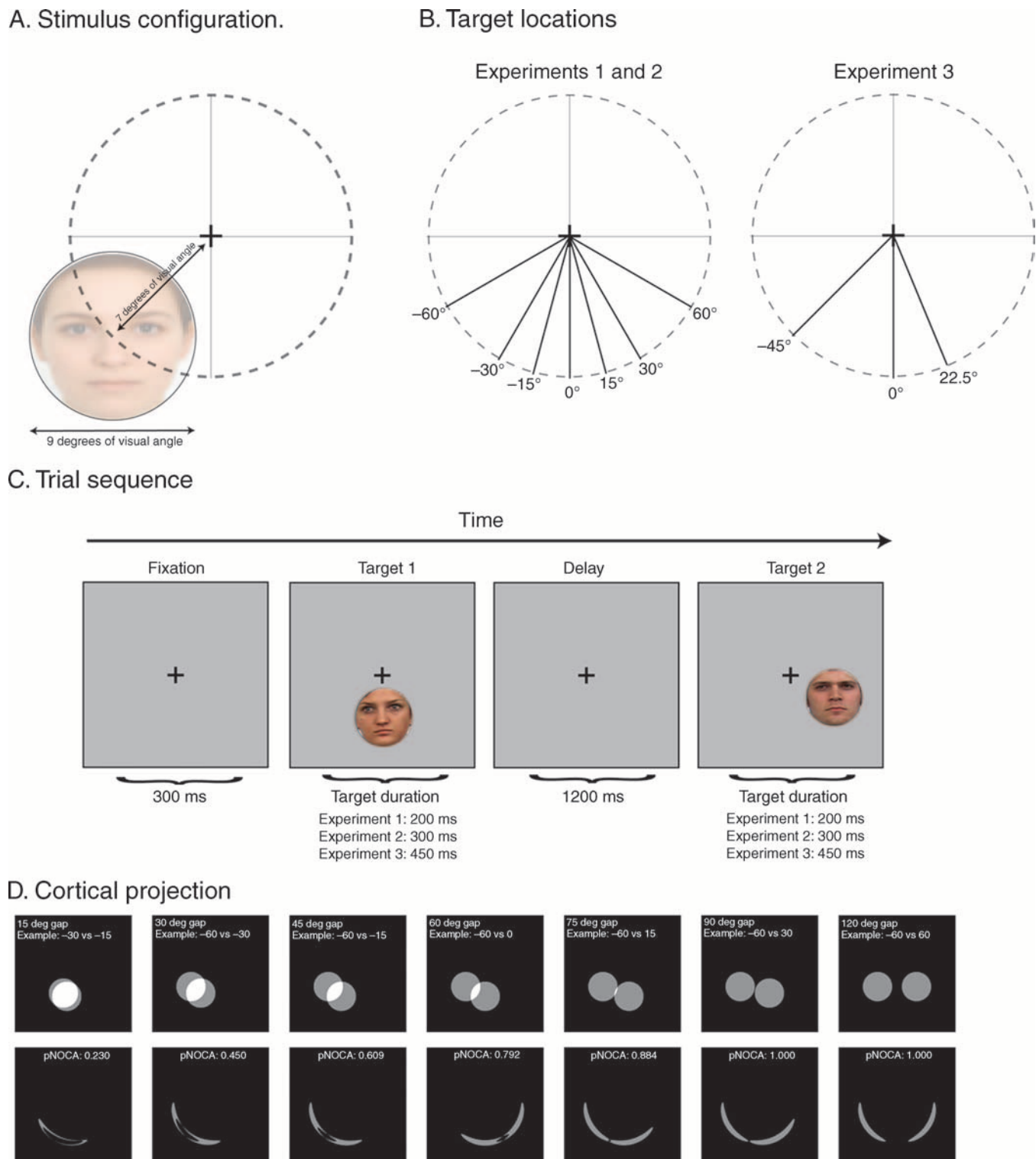
presented; there was a delay period of 1200 ms; the second image was presented; then, the fixation cross was shown again to cue the subject to respond. The duration of the first and second images was fixed within the experiments but varied across experiments (Experiment 1: 200 ms; Experiment 2: 300 ms; Experiment 3: 450 ms). The delay period of 1200 ms was used to allow a sufficient period of time for recovery of the evoked response.

Blocks of trials were randomized and balanced for presentation order and stimulus position. In Experiments 1 and 2, there were seven possible locations for the first target and second target. The counterbalanced set of trials (7 positions [first target] × 7 positions [second target] = 49) was repeated twice for a total of 98 trials per block. In Experiment 1, five participants ran in 6 experimental blocks, and three participants ran in 7 blocks. All of the participants in Experiment 2 ran 6 blocks of trials. After artifact rejection (see below), the number of trials used in the analysis ranged from 542 to 688 and from 500 to 588 for Experiments 1 and 2, respectively. Experiment 3 had four categories (faces, cars, face textures, and car textures). Blocks in Experiment 3 were counterbalanced for presentation order, stimulus position, and stimulus category (3 × 3 × 4 = 36) and repeated twice for a total of 72 trials per block. In Experiment 3, five participants ran in 10 experimental blocks, two ran in 9 blocks, and one participant did 8 blocks. The range of trails used for the analysis in Experiment 3 was 648 to 712 after artifact rejection. Only data from the second target were used in the analysis to avoid differential processing effects between

Figure 1. MEG experiment. (A) Stimulus configuration. Stimuli were 9 degrees of visual angle in diameter, offset 7 degrees of visual angle into the periphery. (B) Target positions. Target position was manipulated by varying the angular offset of the image relative to an arbitrary zero point on the lower horizontal meridian. Experiments 1 and 2 tested six locations: −60, −30, −15, 15, 30, and 60 degrees. Experiment 3 tested three locations: −45, 0, and 22.5 degrees. (C) Trial sequence. Each trial began with a fixation point displayed in the center of the screen. The first target was presented to one location; there was a delay of 1200 ms after the offset of the first target; then, the second target was shown to a second location. The targets were displayed for a time period fixed within experiments but varying across experiments. The subject's task was either to report whether the two targets were of the same gender (Experiments 1 and 2) or whether the two targets were identical images (Experiment 3). (D) Cortical projection. Columns are the unique angular distances between targets used in the analysis. Sample pairs of locations taken from the set of comparisons are shown. The top row shows the overlapping (white) and non-overlapping regions of the visual field (gray) between the two locations. The bottom row is the cortical projection of the non-overlapping region. The pNOCA value (displayed in the graphic) is the proportion of the cortical projection that is non-overlapping relative to the total area that the two stimuli occupy.

the first and second targets (e.g., maintenance of items in short-term memory), although there was little evidence to indicate that this was the case (data not shown).

Experimental task/results

Subjects performed a behavioral task to encourage them to maintain vigilance throughout the duration of the experiment. In Experiments 1 and 2 (face stimuli), participants made "yes" and "no" responses via button press, responding to whether the first and second target images were individuals of the same gender. The mean performance across subjects was 82.6% correct (standard deviation = 8.5%) for Experiment 1 and 86.1% correct (standard deviation = 7.1%) for Experiment 2. The mean reaction times for Experiments 1 and 2 were 797 ms (standard deviation = 260 ms) and 755 ms (standard deviation = 354 ms), respectively. Experiment 3 had four categories of stimuli (faces, cars, scrambled faces, and scrambled cars), three of which were gender neutral. As such, we used an alternative task in which subjects made "yes" and "no" responses to whether the first and second images in each trial were identical. The mean performance across subjects in this task was 93.3% correct (standard deviation = 5.3%), and the mean reaction time was 499 ms (standard deviation = 355 ms).

Eye movements

Eye movements were not recorded during the MEG recording session due to technical limitations. Subjects therefore could have ignored the instructions and fixated the targets. This concern is alleviated by several considerations. First, subjects were trained prior to the MEG recording session and an emphasis was given on maintaining fixation. Second, since the stimuli were presented to random locations, subjects would not have been able to saccade until after the target appeared. If the subject did make a saccade, there would be a large disturbance in the magnetic signal during the trial, and the trial likely would be removed during manual artifact rejection (see below).

MEG recordings and data preprocessing

Neuromagnetic signals were acquired continuously with a 160-channel whole-head axial gradiometer with SQUID-based sensors (KIT, Kanazawa, Japan) sampled at a rate of 1 kHz. Offline, a multitime-shifted principal component analysis denoising algorithm was used to reduce external noise (de Cheveigne & Simon, 2007). Data were resampled to 200 Hz. Trials were epoched into events with a time window extending from 100 ms pre-stimulus onset to 1200 ms post-stimulus onset. Trials were labeled according to the position of the target and the stimulus

category (Experiment 3). Trials were inspected for eye movement artifacts resulting in the rejection of 7.3%, 9.0%, and 8.5% of the trials for Experiments 1, 2, and 3, respectively. A repeated measures ANOVA was conducted to determine if this procedure differentially affected any of the conditions. The analysis indicated that there were no significant differences (Experiment 1, $F(6, 55) = 0.155$, $p = 0.987$; Experiment 2, $F(6, 55) = 0.159$, $p = 0.986$; Experiment 3, $F(2, 23) = 0.009$, $p = 0.991$). After preprocessing, principal component analysis (PCA) was used to reduce the dimensionality of the data set. A fixed criterion of retaining 98% of the variance was used for dimensionality reduction, allowing for a near perfect reconstruction of the data (i.e., little information in the data set is lost through this procedure). This reduced the dimensionality of the data from 157 recording channels to, on average, 76 spatial components.

Classification analysis

Linear discriminant analysis (Duda, Hart, & Stork, 2001) was used to perform single trial classification. Classifiers were trained and tested on trial scores from the PCA. Separate classifiers were trained on each time point in the time series. With the data resampled to 200 Hz, the effective resolution of the analysis was 5 ms. The data were analyzed by training the classifier to recover the position of the stimulus for all possible pairwise combinations of locations (e.g., −60 and −15 degree offsets; see Table 1). In Experiments 1 and 2, stimuli were presented to seven locations resulting in 21 comparisons. In Experiment 3, the stimuli were presented to three locations with irregular spacing resulting in three pairwise comparisons. A separate analysis was performed for object category in Experiment 3 using the same procedure. There were six

possible pairwise comparisons for the four stimulus categories, but only four comparisons were used. Two of the comparisons were excluded (faces vs. scrambled cars and cars vs. scrambled faces) from the analysis as they were of no theoretical interest.

Cross-validation

The training and testing of the classifier was performed on separate data. Trials from each condition were pooled by stimulus position (and category for Experiment 3) across runs. The pooled data set was randomly divided into 20 subsets. These subsets were used to perform a *k*-fold cross-validation with a ratio of 19:1 training to test. In this procedure, one subset was removed (test data) and the remaining 19 subsets were pooled and used to train the classifier. After training, performance was evaluated by testing individual trials from the test data. The procedure was repeated 20 times, each time leaving out one of the subsets such that each trial was tested once.

Modeling the decoding performance for position

We created a model that assumes topographic organization to examine the relationship between activation patterns in cortical maps and decoding performance. The analysis uses single time points; classification is thus based on the difference between the magnetic signal topographies for the two locations used in a given comparison. Signal topographies are determined by activation patterns in cortical maps, which are determined by the retinal projection of the stimuli. Seven unique angular distances were tested in the set of pairwise comparisons for Experiments 1 and 2

Experiment	Angular distance between targets (degrees)	Number of comparisons	Listed comparisons	Proportion non-overlapping cortical area (pNOCA)
1, 2	15	4	(−30, −15); (−15, 0); (0, 15); (15, 30)	0.23
1, 2	30	5	(−60, −30); (−30, 0); (−15, 15); (0, 30); (30, 60)	0.45
1, 2	45	4	(−60, −15); (−30, 15); (−15, 30); (15, 60)	0.609
1, 2	60	3	(−60, 0); (−30, 30); (0, 60)	0.792
1, 2	75	2	(−60, 15); (−15, 60)	0.884
1, 2	90*	2	(−60, 30); (−30, 60)	1
1, 2	120*	1	(−60, 60)	1
3	22.5	1	(0, 22.5)	0.343
3	45	1	(−45, 0)	0.609
3	67.5	1	(−45, 22.5)	0.838

Table 1. Angular distance. The table shows the unique angular gaps between target experiments. The first column indicates the experiment (1, 2, or 3). The second column is the angular gap between targets used in the comparison. The third column indicates the number of comparisons for each angular gap size. The fourth column is the listed comparisons. The fifth column is the proportion of space that is non-overlapping in the cortical projection (pNOCA; see text for description). *Notes*: *For 90 and 120 degrees, the stimuli are completely non-overlapping and have identical pNOCA values (1.0).

and three distances in Experiment 3 (Table 1). Figure 1D (top row) graphically shows the area of the visual field occupied by the stimuli for each angular gap tested in Experiments 1 and 2. The white region in the figure is the overlapping region of the visual field occupied by the two stimulated locations. This shared region is the same for the two conditions and thus will not evoke a differential cortical response (ignoring image variations). The non-overlapping region, shown in gray, is spatial difference between the conditions that will determine classification performance. Since the MEG signal reflects cortical activity, not retinal, a log polar transformation was used to approximate the cortical projection (Schwartz, 1977) of the non-overlapping region (Figure 1D, bottom row). The absolute size of this region (i.e., in units of mm^2) cannot be determined, since the source of the MEG signal cannot be attributed to single or multiple cortical generators (e.g., V1, V2) with certainty. We thus created a metric (proportion non-overlapping cortical area, pNOCA) that is the area of the non-overlapping region scaled by the total area that the stimuli independently occupy. Classification performance is expected to scale with pNOCA (i.e., the larger the non-overlapping area, the better the classifier will perform), assuming a topographical organization. This expected relationship is based on a simplified view of the source of the MEG signal and cortical representation. For example, the model does not take into account the fact that MEG signal reflects activity of multiple areas nor does it account for the distance between the wedges of the non-overlapping regions, the folding of the cortex, and that a log polar transformation is only an approximation of the retinal to cortical mapping (Schira, Tyler, Spehar, & Breakspear, 2010).

Results

Decoding object position from neuromagnetic recordings

A linear classifier was trained to recover the position of the stimulus from the MEG recordings. The analysis was conducted for each point in the time series (sampled at 200 Hz) to obtain an accurate measure of time course. Figure 2A shows the performance of the classifier, averaged across all comparisons and across subjects. Figure 2B shows the data for Experiment 2. In interpreting the results, it is important to note that peaks in performance (e.g., at 115 ms) are different from peaks or components in traditional EEG and MEG analyses. Here, peaks are indicative of reliable differences in the MEG signal topography between conditions that can be used by the classifier to decode the stimulus, which may or may not correspond to peaks in the evoked response. We arbitrarily defined the onset as the first three consecutive points above chance at

a threshold of $p < 0.01$ (uncorrected). Using this criterion, the onset was 80 ms and 60 ms for decoding stimulus position for Experiments 1 and 2, respectively. After onset, performance rises to a peak at 115 ms, and then decays slowly with an intermittent second peak (discussed below). Decoding performance returns to chance at approximately 1000 ms post-stimulus onset in Experiment 1 and 600 ms post-stimulus onset in Experiment 2.

Topographical organization and decoding performance

The representation of the visual field is encoded in topographic maps in visual areas and other areas throughout the brain (DeYoe et al., 1996; Engel et al., 1997; Hagler & Sereno, 2006; Kastner et al., 2007; Sereno et al., 1995; Sereno, Pitzalis, & Martinez, 2001; Silver & Kastner, 2009; Wandell et al., 2007). We next examined if topographic organization could account for classification performance. To this end, we calculated a metric pNOCA that is derived from a model that assumes topographic organization (see Methods and materials section) and is expected to be proportional to classification performance. We first tested the model with data from 115 ms post-stimulus onset, a time when there was a highly reliable difference between the conditions (as indicated by good performance) and the source of the MEG signal has previously been localized to primary visual cortex and extrastriate visual areas (Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002). Figure 2C plots performance as a function of pNOCA for the data acquired at 115 ms. As can be seen in the figure, the relationship between decoding

Figure 2. Position classification results for Experiments 1 and 2. (A, B) Classification performance for spatial position. Plots show the performance of the classifier for pairwise decoding of target position, evaluated as a function of time. Mean performance is indicated by a thick blue line; the gray shaded region between the thin blue lines is ±1 standard error of the mean (*SEM*) across subjects. Chance performance is 0.5 correct. Blue stars plotted at 0.4 (*y*-axis) indicate above chance classification performance ($p < 0.01$, uncorrected). Vertical dashed lines indicate the onset and offset of the stimulus. Arrows denote a second peak in classification performance located at approximately 320 ms (Experiment 1) and 415 ms (Experiment 2). (C, D) Non-overlapping region and classification performance for the first peak. Plots show the performance of the classifier as a function of pNOCA for data acquired at 115 ms post-stimulus onset for Experiments 1 and 2, respectively. Errors bars are *SEM*. The line is the fit from the regression analysis (Experiment 1: slope = 0.30, *y* intercept = 0.55 proportion correct; Experiment 2: slope = 0.31, *y* intercept = 0.54 proportion correct). (E, F) Non-overlapping region and classification performance across the time series. Plots show the performance of the classifier binned according to the pNOCA value. Red stars plotted at 0.4 (*y*-axis) indicate a significant correlation ($p < 0.01$, uncorrected) between classification performance and pNOCA.

performance and pNOCA is highly linear. The results of a regression analysis confirm that there is a close coupling between performance and pNOCA in both experiments (Experiment 1: $R^2 = 0.43$, $F(1, 167) = 126.80$, $p < 0.0001$;

Experiment 1: $R^2 = 0.42$, $F(1, 167) = 122.63$, $p < 0.0001$). Thus, topographic organization accounts well for performance at 115 ms post-stimulus onset. We subsequently expanded the analysis to the time series by correlating
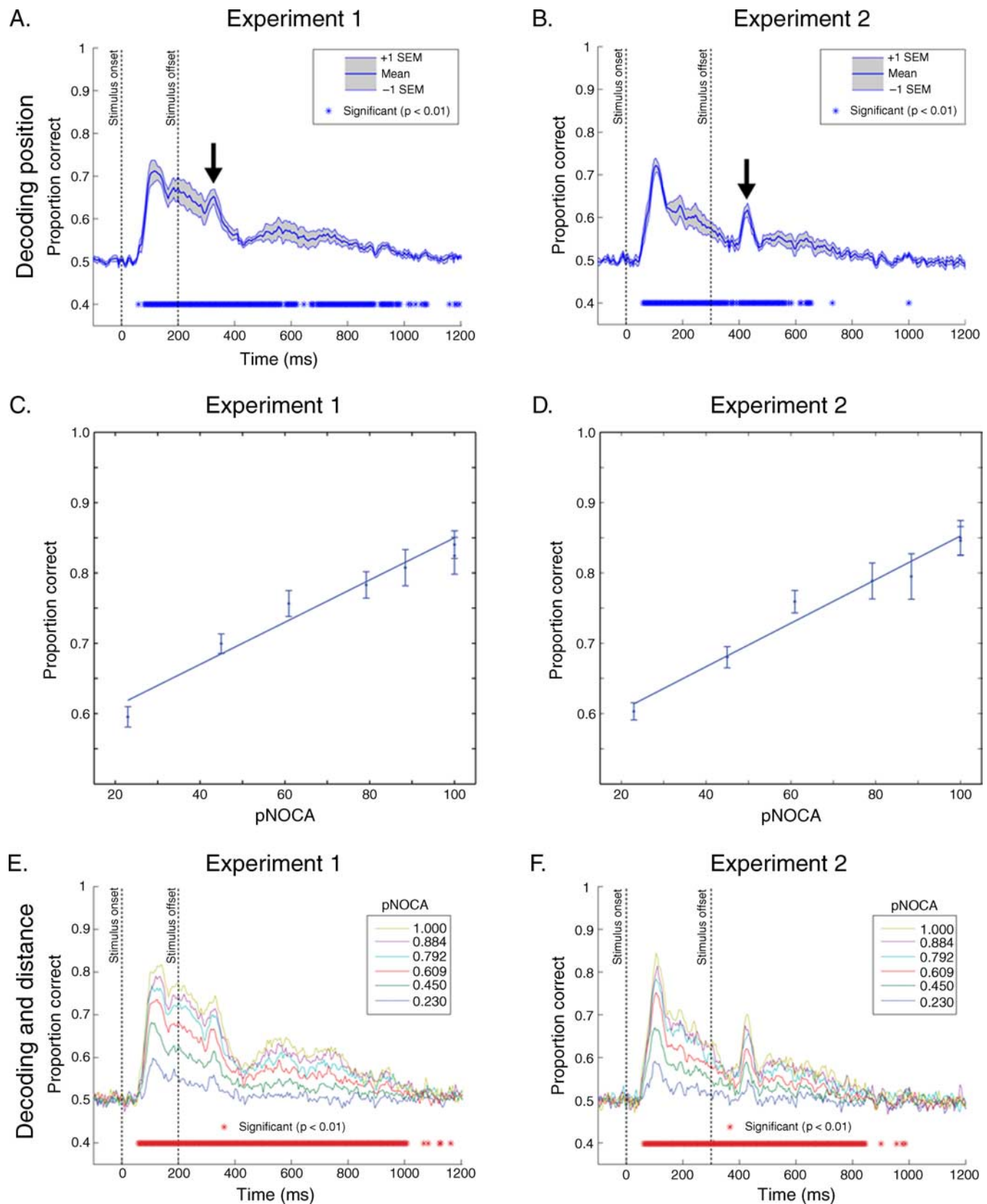
pNOCA with decoding performance for each time point. The plots shown in Figures 2E and 2F demonstrate the strong relationship between pNOCA and performance for Experiments 1 and 2, respectively. Across virtually the entire time series, higher pNOCA values correspond to better classification performance. The correlation analysis confirms this trend showing that the correlation between pNOCA and performance was significant for nearly the entire period of time performance was above chance (Figures 2A and 2B).

Second peak in decoding performance

There is a prominent second peak (denoted by the arrows in Figures 2A and 2B) at 320 ms in Experiment 1 and at 420 ms in Experiment 2. This second peak is well predicted by the timing of the offset (Experiment 1, 200 ms; Experiment 2, 300 ms) and the lag in the peak observed for the onset of the stimulus (115 ms). Together, these numbers sum (stimulus offset + lag) to the approximate location of the second peak (Experiment 1, 315 ms; Experiment 2, 415 ms). This implicates the source of the peak to be a response to the stimulus offset.

## Object category and the representation of position

In Experiment 3, we examined the coding of position and category. The number of locations was reduced to three (-45, 0, and 22.5 degree offsets). In each location, participants were shown four categories of stimuli (faces, cars, face textures, car textures). The duration of the targets was also extended to 450 ms to further examine the offset response. Replicating the results of the first two experiments, the classifier could accurately recover the position of the target; decoding performance systematically varied with the distance between the targets used in the comparison (Figure 3A); and there was a prominent second peak in performance that shifted in correspondence with the offset (demarked by the arrow in the figure).

We next trained the classifier to recover the category of the stimulus (face, car, face texture, car texture). Figure 3B shows the outcome for four comparisons; of which, the category of the stimulus could be recovered for three of the four. Using the same criterion described above (three consecutive points above chance, thresholded at $p < 0.01$, uncorrected), the onset latencies for the three comparisons were 105 ms, 110 ms, and 135 ms for faces vs. face texture, car vs. car texture, and face vs. car comparisons, respectively. Performance of the classifier was above chance until at least 650 ms for these three comparisons, an interval that overlapped with the offset response. In the comparison between the two textures (car textures and face textures), performance of the classifier was only sporadically above chance, likely reflecting the effect of

multiple comparisons, but never reached our threshold of three consecutive points above chance. This comparison is notably important because the images were generated with an algorithm that replicates image statistics (Portilla & Simoncelli, 2000). This failure of the classifier to recover the category of the texture shows that decoding for the comparisons between faces and cars (i.e., object category decoding) cannot be accounted for by differences in low-level image statistics (e.g., contrast, spatial frequency; see Honey, Kirchner, & VanRullen, 2008).

In the category analysis, the classifier was trained and tested using only category labels (e.g., face vs. car). Since the position of the stimulus was not labeled in the training, the classifier was effectively blind to the variation in position. Accurate decoding therefore would require the classifier to learn to generalize category information across retinal locations, i.e., position-invariant classification. To examine this further, we trained a new set of classifiers to recover category using two different procedures (Figure 3C). The difference between the procedures was that in one case the classifier could use activation patterns that were specific to a location in the visual field. In the first procedure, the classifier was allowed to use location-specific information by training and testing the classifier using data from the same location. Three classifiers were trained,

Figure 3. Position and category. Plots show the average performance of the classifier for recovering the target position and stimulus category. Mean performance is indicated by a thick line; the shaded region is ± 1 standard error of the mean (*SEM*) across subjects. (A) Classification performance for spatial position. Plots show the results of the classification analysis for the three pairwise comparisons: -45 vs. 22.5 degree offset, -45 vs. 0 degree offset, and 0 vs. 22.5 degree offset. Vertical dashed lines indicate the onset and offset of the stimulus. Arrow denotes the second peak in classification performance. Blue stars plotted at 0.4 (*y*-axis) indicate a significant correlation ($p < 0.01$, uncorrected) between performance and pNOCA. (B) Classification performance for stimulus category. Plots show stimulus category classification for the classifier trained using data from all three locations. Individual plots display the four pairwise comparisons for stimulus category: faces vs. scrambled faces, faces vs. cars, cars vs. scrambled cars, and scrambled faces vs. scrambled cars. Blue stars plotted at 0.4 (*y*-axis) indicate above chance classification performance ($p < 0.01$, uncorrected). Note that the scale for the *y*-axis has changed. (C) Specific location and novel location classifier training. The position-specific classifier was trained and tested at the same location in the visual field. The position novel classifier was trained at two locations and tested at the third untrained location. (D) Comparison between location-specific and novel location classification. Plots show the average performance across locations for the position-specific (red) and position novel (blue) classifiers. Individual plots display the three pairwise category comparisons. Blue stars plotted at 0.4 (*y*-axis) indicate a significant difference in the performance between the two classifiers ($p < 0.01$, uncorrected).
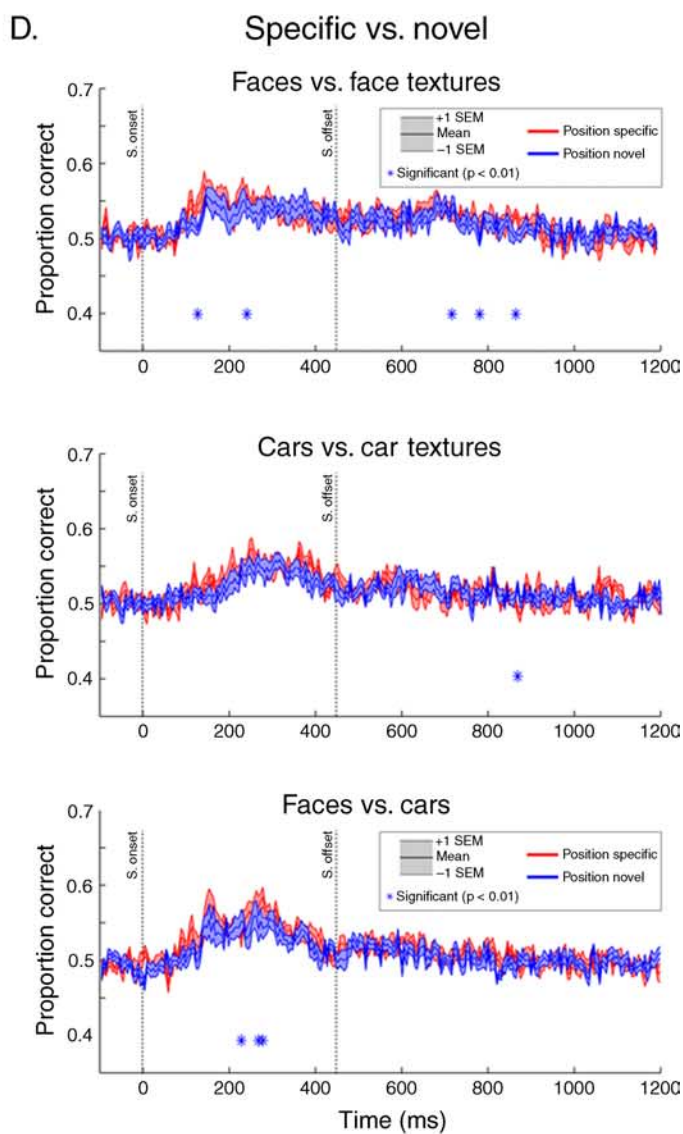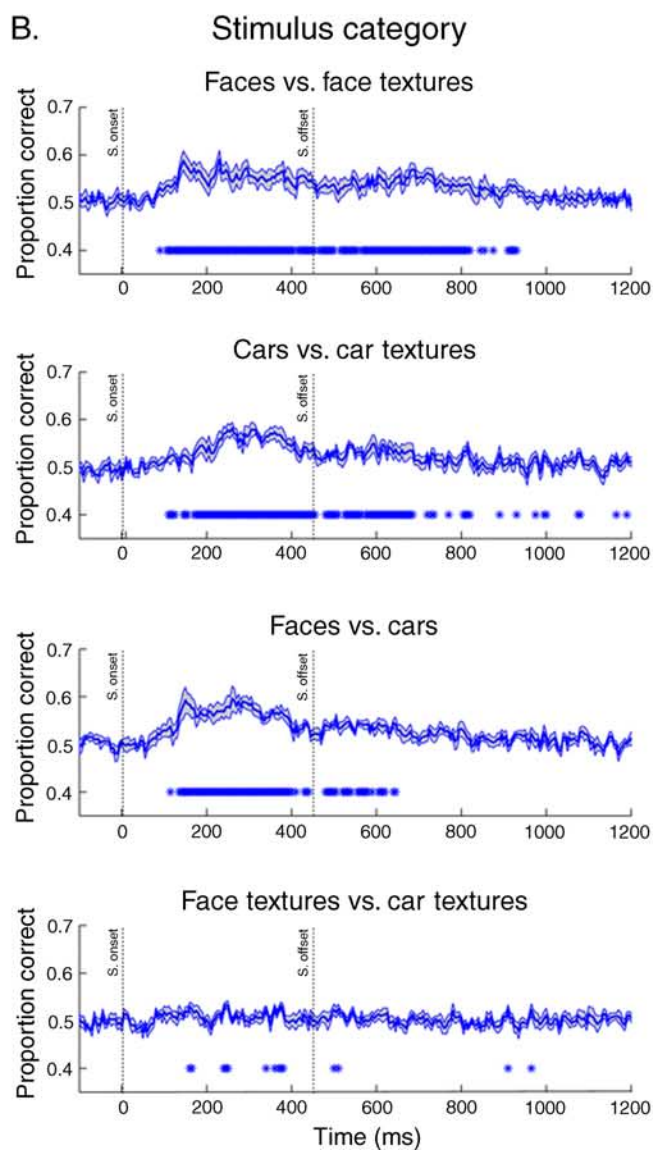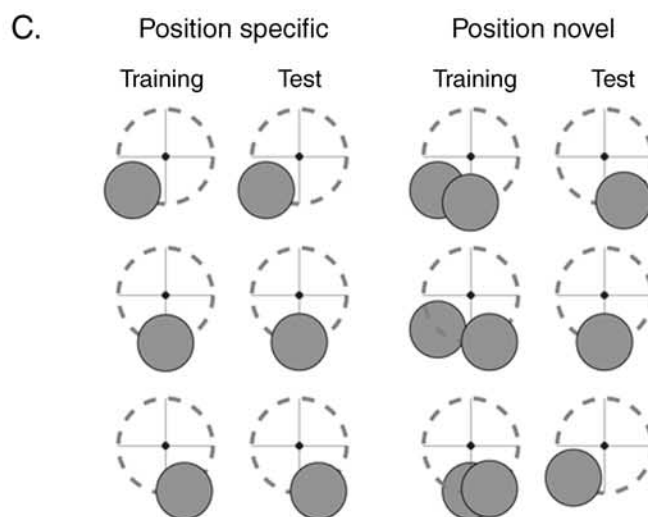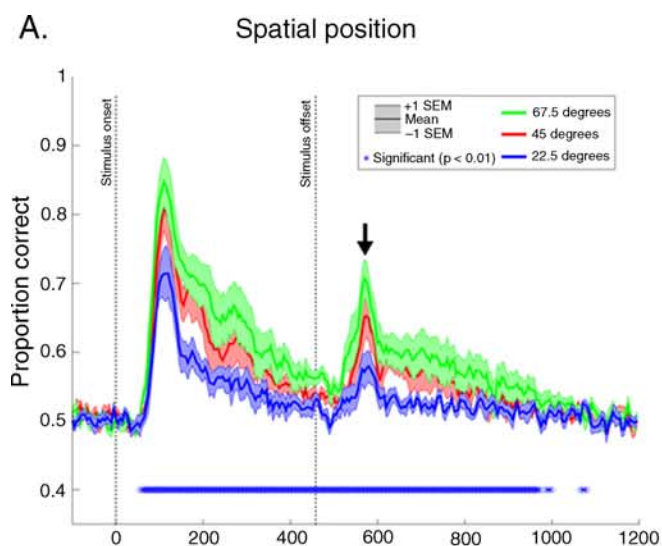
each specialized for one location in the visual field (-45, 0, and -22.5 degree offsets). In the second procedure, the classifier was trained to recover the category from a novel location, not included in the training. The classifier was trained using data from two locations and then tested with data from the third location. Here, the classifier exclusively must rely on position-invariant responses. Three comparisons (faces and face textures, cars and car textures, and faces and cars) were analyzed using these two procedures. Figure 3D shows the results. As can be seen in the figure, the time course is virtually identical to our initial analysis (shown in Figure 3B). Importantly, there was little observable benefit to training and testing at the same location over testing at an untrained location, which indicates

that the classifier was relying on location-invariant brain activity to recover the category of the stimulus. Taken together with the outcome of the previous analysis, these findings indicate that position-invariant object information is extracted by the visual system as early as 105 ms for the distinction between objects and non-objects, represented by the object texture comparison, and 135 ms for distinguishing between object categories (face vs. car).

Discriminant cross-training (DCT)

We observed a second peak in classification performance for retinal position that corresponds with the offset
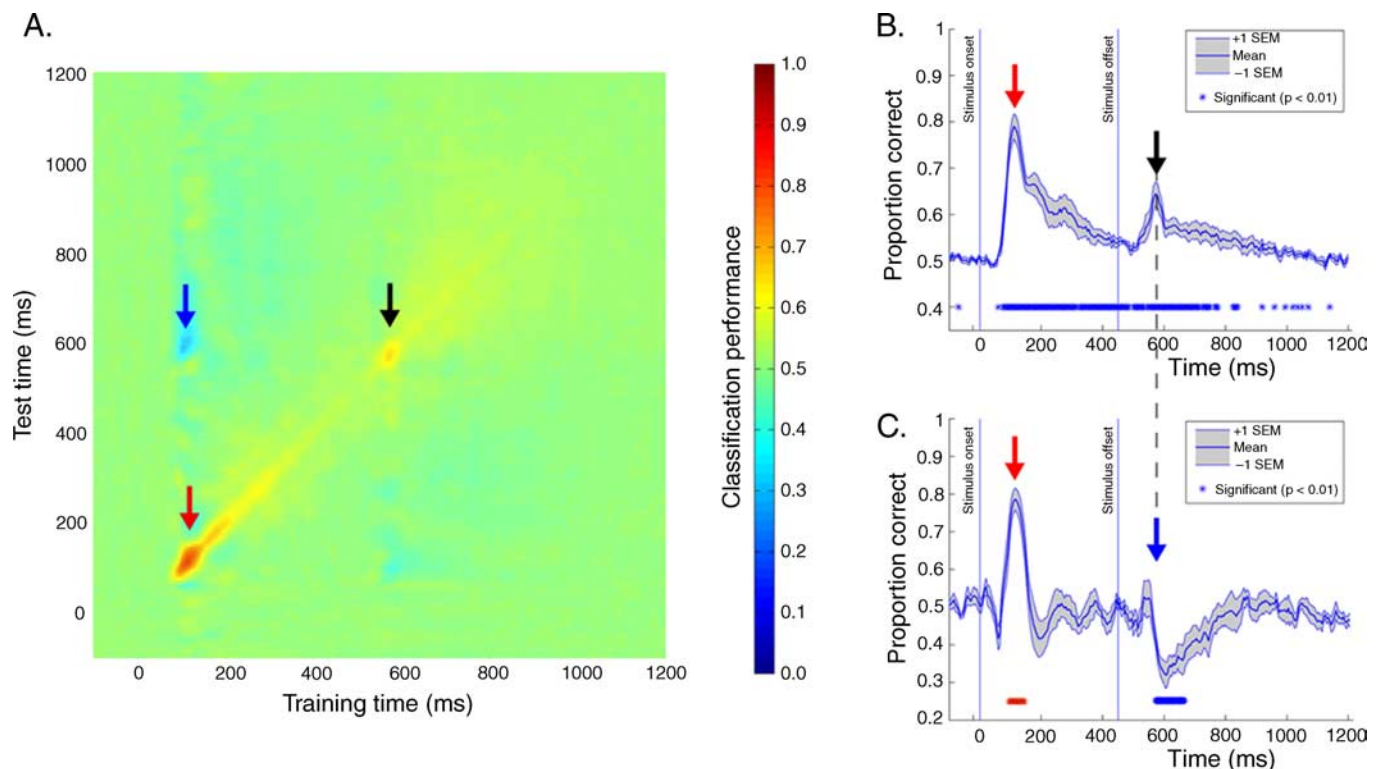

Figure 4. Discriminant cross-training. (A) Results of discriminant cross-training analysis. The image shows the average performance of the classifier for decoding stimulus location as a function of training time (*x*-axis) and test time (*y*-axis). Arrows superimposed on the image denote three notable features. The red arrow located on the identity axis marks the initial peak in classification performance located at approximately 115 ms training and 115 ms test (see also Figure 3A). The black arrow also located on the identity axis marks the second peak in classification performance located at approximately 570 ms training and 570 ms test, corresponding to the offset of the stimulus (see also Figure 3A). The blue arrow located at approximately 115 ms training and 570 ms test indicates transfer from training on the initial peak to testing on the second peak, albeit in the form of below chance performance. (B) Identity training. This plot shows the performance of classifier extracted from the identity axis of the discriminant cross-training matrix. This is equivalent to the previous time series analysis using a moving time window with training and test performed on the same time points. The red and black arrows mark the first and second peaks in classification performance, respectively. Mean performance is indicated by a thick blue line; the gray shaded region between the thin blue lines is ±1 standard error of the mean (*SEM*) across subjects. Blue stars plotted at 0.4 (*y*-axis) indicate above chance classification performance ($p < 0.01$, uncorrected). (C) Performance of classifier trained at 115 ms. Plot shows the performance of classifier when training is performed at 115 ms post-stimulus onset. Expectedly, classification performance peaks at 115 ms (marked with the red arrow), as this is the period of time during which the classifier was trained. Notably, classification performance falls significantly below chance at approximately 570 ms post-stimulus onset (marked with the blue arrow). Mean performance is indicated by a thick blue line; the gray shaded region between the thin blue lines is ±1 standard error of the mean (*SEM*) across subjects. Stars plotted at 0.4 (*y*-axis) indicate significantly above (red) or below (blue) chance classification performance ($p < 0.01$, uncorrected).

of the stimulus. The second peak is indicative of a reactivation of visual areas in response to the offset. To examine this further, we developed a procedure that we refer to as discriminant cross-training (DCT). Akin to cross-training procedures used in fMRI (Polyn, Natu, Cohen, & Norman, 2005; Stokes, Thompson, Cusack, & Duncan, 2009), DCT aims to access whether two representations are similar (Kriegeskorte, 2011; Norman, Polyn, Detre, & Haxby, 2006). The logic of DCT is that a trained discriminant can act as a virtual electrode in time. In training, the classifier weights inputs to maximize the difference between conditions for testing. The weighting will reflect cortical activity that is useful for the classification and will change based on the time of the training. For example, early in the time series, early visual areas are encoding the stimulus. The classifier will accordingly give greater weight to signals originating from the visual cortex to maximize decoding performance. The DCT analysis presupposes that if these same areas become active later in the time series, the classifier will generalize at this later time.

To test this, we trained separate classifiers to recover the position of the stimulus for each time point, and then tested whether these trained classifiers generalize to the data from other time points across the entire time series. The generalization of classifiers is summarized as a matrix (Figure 4A) in which columns correspond to the times points on which the classifier was trained, and rows correspond to the points on which the classifier was tested. The identity of the matrix (the diagonal) is equivalent to the analysis conducted previously (i.e., the classifier was trained and tested at the same time point; Figure 4B). Using this cross-discriminant analysis, we sought to address the relationship between the first and second performance peaks (Figure 4C). Specifically, we examined the generalization of the classifier trained at the first peak across the entire time series. This analysis showed that the classifier generalized only for a narrow time window of 60 ms surrounding the time of the training and returned to chance until the offset response. However, the chance performance outside this temporally specific window does not imply that positional information is absent during this period, because the stimulus position was successfully predicted when classifiers were trained and tested for each time point separately (Figure 4B). The temporal specificity observed here suggests that the activity pattern predicting the stimulus position changes continuously as a wave of evoked activity propagates through the visual hierarchy. In support of this notion, classifiers trained at other fixed time points also exhibited narrowly tuned temporal specificity (see Supplementary Figure 1). At the time of the offset, the classifier trained at the first peak began to generalize again (Figure 4B). However, contrary to our expectation, the generalization came in the form of systematic errors resulting in below chance performance (denoted by the blue arrow in Figure 4B; also visibly apparent in Figure 4A). The temporal specificity and generalization from the onset to the offset support our

presupposition that cross-discriminant analysis can act as a virtual electrode to reveal the consistency of informative activity patterns across different time points.

Discussion

We used linear discriminant analysis to recover the position and category of stimuli from MEG recordings. From single trials, both the position and category could reliably be decoded. The classification performance for position was in good agreement with topographical organization of the visual field in the cortex, and the onset of above chance performance (~70 ms) was concordant with previous estimates of the time visually evoked response, thus validating the decoding approach as a sensitive measure. We then showed that intact objects could be dissociated from textures with similar image statistics as early as 110 ms after stimulus onset, while decoding two high-level categories of stimuli (faces vs. cars) required slightly more time (135 ms). Secondary analysis further revealed that category decoding was based on location-invariant responses to the stimulus.

Decoding position in the visual field

Neurophysiological responses to manipulations of position have been studied extensively using EEG and MEG. These studies generally sought to address issues of source localization (recent examples include Ales, Carney, & Klein, 2010; Hagler et al., 2009; Im, Gururajan, Zhang, Chen, & He, 2007; Poghosyan & Ioannides, 2007) as topographic organization in visual areas provides an excellent reference for validation. In the present study, we similarly took advantage of topographic organization to validate our approach. We found performance of the classifier to be compatible with the established timing and topographic organization in visual areas. The classifier could decode position as early as 60 ms post-stimulus onset, which is consistent with recorded responses in early visual areas (Aine, Supek, & George, 1995; Brecelj, Kakigi, Koyama, & Hoshiyama, 1998; Di Russo et al., 2002; Jeffreys & Axford, 1972; Nakamura et al., 1997; Portin et al., 1999; Supek et al., 1999). The performance of the classifier was well described by topographic organization in the visual cortex (DeYoe et al., 1996; Engel et al., 1997; Sereno et al., 1995, 2001). Stimulating different retinal locations would elicit predictable activation patterns in topographic maps of the visual field, which in turn would produce distinct magnetic signal topographies. We constructed a simple model of the expected differences in the evoked response and found that the model metric (pNOCA) was proportional to classification performance. The congruence of classification performance and established properties of early visual

areas validates the decoding approach as a sensitive measure for MEG research and presumably EEG as well.

Onsets and offsets in the visual field

In all three experiments, the offset of the stimulus evoked a response that was sufficient to recover the position of the stimulus. Visual offsets are of notable interest because they reveal the time that a stimulus is encoded after it is no longer physically present. In our experiments, we found that the position is represented for up to 500 ms after it is physically removed (Experiments 1, 2, and 3). Similar to the onset response, the performance of the classifier was well described by the topographic model. The robust response to offset suggests that offsets may have an understated role in perception and cognition.

Visual onsets and offsets have been studied with single unit, EEG, and MEG (Kouijzer et al., 1985; Parker, Salzen, & Lishman, 1982; Phillips & Singer, 1974). Consistent with early MEG studies, we found that onsets and offsets elicit distinct responses. We further observed an intriguing relationship between the onset and offset responses. The DCT analysis showed that a classifier trained on the onset was able to generalize to the offset but in the form of significantly below chance performance. This inversion can be accounted for by anticorrelated onset and offset MEG signal topographies. Figure 5A shows the three stimulus locations tested in Experiment 3 with their respective field topographies at 115 ms post-stimulus onset. The third column in the figure shows the average correlation between individual subjects' onset topography at 115 ms and offset topographies as a function of time from the offset. Note that the activation pattern becomes negatively correlated around 100 ms. Figure 5B shows the impact of the anticorrelated onset and offset on classification performance (see also Figure 4). In the figure, the onset and offset (aligned at zero in the figure) are similar for the first 80 ms at which point the two curves diverge. Onset performance then rises sharply. Performance for the offset, in contrast, remains at chance until 100 ms and then falls below chance. This time corresponds with the time that the onset and offset magnetic field topographies become anticorrelated (Figure 5A), providing an account for the below chance decoding performance. One explanation for the anticorrelated offset is that neurons that respond to the stimulus briefly drop below baseline firing rates after stimulus offset. Any cell that responds primarily to stimulus changes (i.e., with temporal high-pass response characteristics) might be expected to show this overshoot upon stimulus offset. The pattern of activity in such high-pass cells after stimulus offset would therefore briefly be opposite to the activity evoked by stimulus onset. As an interesting aside, a similar polarity reversal has been observed in response to auditory transitions (Chait, Poeppel, de Cheveigne, & Simon, 2007). While the aims of that study were different from the current one, it is notable that the polarity reversals in both studies arose in the context of responses to temporal edges.

The prominence of the offset response is indicative of functional relevance. Studies examining the role of transients in object recognition and memory have mainly
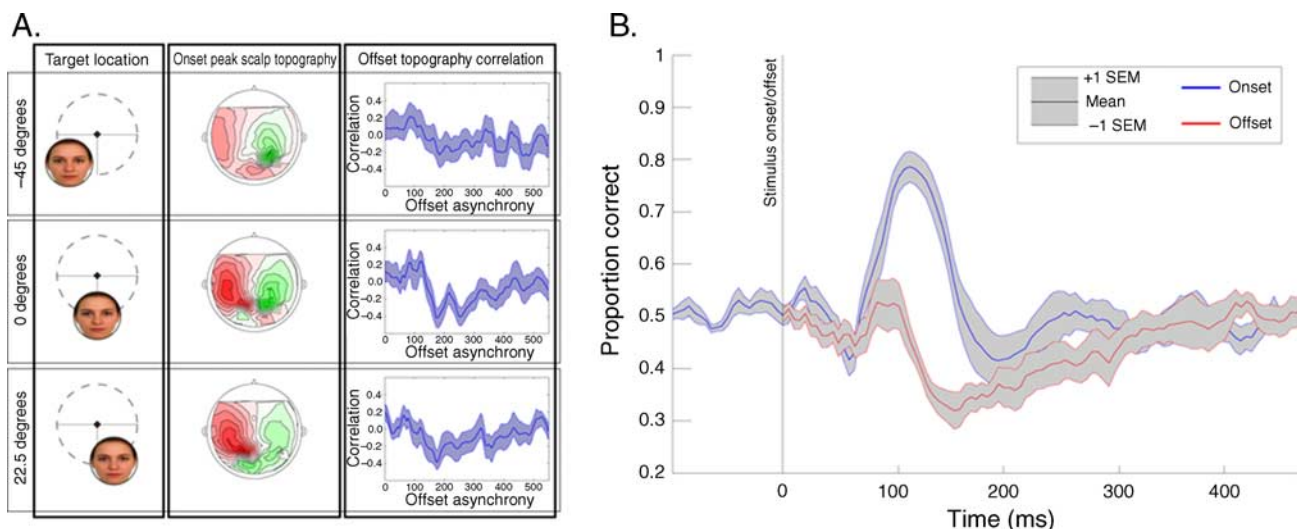

Figure 5. Onset and offset response. (A) Anticorrelated scalp topographies (Experiment 3). The target locations (first column) and the respective scalp topography for the onset response at 115 ms (second column) from a representative subject are shown. The third column shows the average correlation across subjects between the magnetic signal topography for the onset response at 115 ms and the magnetic signal topography for the offset. The shaded region is ±1 standard error of the mean (SEM). (B) Classification performance for onsets and offsets. The data from Experiment 3 when the classifier was trained at 115 ms post-stimulus onset are replotted. Classification performance for the onset (shown in blue) and offset (shown in red) are aligned at time zero to the stimulus onset and offset for visualization. The shaded region is ±1 standard error of the mean (SEM) across subjects.

focused on onsets. Onsets signal the introduction of new, possibly behaviorally relevant information (e.g., a predator) and accordingly capture attention (Gellatly & Cole, 2000; Gellatly, Cole, & Blurton, 1999; Jonides, 1981; Jonides & Yantis, 1988; Theeuwes, 1991). Offsets instead are indicative of the removal of a stimulus from the visual field. It may be evolutionarily advantageous to temporarily store information removed from online visual processing. Offsets conceivably could serve as a trigger for the visual system to store information in memory. Consistent with this, studies have found that visual short-term memory traces are stronger for offsets (Cole, Kentridge, Gellatly, & Heywood, 2003; Mondy & Coltheart, 2000). Our findings provide further support for this in showing that behaviorally relevant visual information is embedded in the offset response. In addition to position, the offset encodes information about category. In the comparison between faces and cars, decoding performance falls to chance before the offset of the stimulus and then rises again to above chance levels after the offset. This recovery following the offset indicates that the offset response encodes the category of an object that has been removed from online visual processing, which could be used by the visual system as an input for a short-term memory trace (Sperling, 1960).

Decoding invariant object category information

In agreement with neurophysiological studies showing sensitivity of IT neurons to stimulus position (DiCarlo & Maunsell, 2003; Op De Beeck & Vogels, 2000), human imaging studies have found that areas in human ventral visual cortex are topographically organized (Arcaro, McMains, Singer, & Kastner, 2009; Brewer, Liu, Wade, & Wandell, 2005; Sayres & Grill-Spector, 2008) and that the position of an object can be decoded using multivoxel pattern analyses in higher areas such as LO (Carlson et al., 2011; Cichy et al., 2011; Sayres & Grill-Spector, 2008; Schwarzlose et al., 2008). Similarly, MEG studies using source localization have shown that varying the position of a stimulus evokes distinct patterns of activity in ventral temporal cortex (Liu & Ioannides, 2006, 2010). Solidifying position coding in ventral temporal cortex introduces an important constraint on models of object recognition, as these models need to account for the sensitivity of IT neurons to retinal location. One possibility is that position invariance can be addressed at the population level by integrating responses across individual neurons (DiCarlo & Cox, 2007; for a critique of this account, see Robbe & Op de Beeck, 2009). This account has received empirical support from both neurophysiological and imaging studies showing that linear classifiers can read out category and identity information from areas in the ventral temporal pathways across changes in position and size (Carlson et al.,

2011; Cichy et al., 2011; Hung et al., 2005; MacEvoy & Epstein, 2007; Sayres & Grill-Spector, 2008; Schwarzlose et al., 2008; Williams et al., 2008).

We found no observable benefit to training and testing on the same location over testing on untrained locations (Figure 3D). This result seemingly contradicts the findings of neuroimaging studies that have shown a performance benefit for location-specific decoding (Cichy et al., 2011; Kravitz et al., 2010; Sayres & Grill-Spector, 2008; Schwarzlose et al., 2008) and the theoretical proposal that object representations are location specific (Kravitz et al., 2010). Given the replicability of this result, the failure to find a location-specific benefit suggests that MEG recordings may lack the signal to noise to detect fine-grained location-specific object representations. Compatible with this, fMRI studies examining the encoding of the object properties of pose and position necessitated the use of more sensitive paradigms and analyses, i.e., fMRI adaption (Grill-Spector et al., 1999) and multivoxel pattern analysis. The inferior spatial resolution of MEG thus may ultimately limit the technique's ability to detect subtle activation patterns associated with location-specific object representations, even using sensitive pattern classification techniques. Cichy et al. (2011) recently have shown that object representations in LOC exist both as position-specific (as indicated by a location-specific training benefit) and broader activation patterns that are position tolerant (as indicated by the capacity to decode object category and exemplar across locations) using fMRI. The absence of location-specific benefit and the capacity to decode category from untrained locations indicate that decoding in the present study relies on a neural activity from a representation that is position tolerant, possibly as high as the semantic level, which would be concordant with the finding that we could be not discriminate between face and car textures. This "high-level" account would interestingly also suggest that semantic categorization occurs as quickly as 110 ms (Thorpe, Fize, & Marlot, 1996). Future research could investigate this possibility using artificial stimuli.

Our findings place an upper bound on the time necessary for the brain to construct an object representation invariant (or sufficiently tolerant) to position, which is a useful constraint on models of object recognition. Generally, short latencies favor feed-forward models (Fukushima, 1980; Riesenhuber & Poggio, 1999; Rolls, 1991; Serre et al., 2007; VanRullen & Thorpe, 2002), while longer latencies are indicative of recurrent process models (Bullier, 2001; Hochstein & Ahissar, 2002). In addressing the issue of latency, neurophysiological data have been mixed, and there is very limited human data. Monkey neurophysiological studies have found that invariant category and identity information can be read out as early as 100 ms (Hung et al., 2005) or as late as 300 ms (Freiwald & Tsao, 2010), and intracranial recordings in humans have found invariant object representations in

population codes as early as 100 ms post-stimulus onset. Here, we found that "objectness" (objects vs. images controlled for image statistics) can be decoded as early as 110 ms and category (faces vs. cars) can be decoded as early as 135 ms. Our findings therefore show that the visual system rapidly recovers location-invariant object category information, which arguably favors feed-forward models. At a minimum, this short latency constrains the number of computational steps the brain uses to derive a representation of an object that is invariant to position. There are, however, several issues to consider in interpreting the data across these studies. There is the animal model (humans vs. non-human primates), the level of analysis (category vs. exemplar level decoding), methods used for the measurement of neural activity, and the types of variations examined (position, size, viewpoint). More systematic study of these factors will be necessary to elucidate how we so easily recognize objects—in just a fraction of a second—across the highly variable circumstances that we encounter objects in our environment.

Acknowledgments

References

Aine, C. J., Supek, S., & George, J. S. (1995). Temporal dynamics of visual-evoked neuromagnetic sources: Effects of stimulus parameters and selective attention. *International Journal of Neuroscience, 80,* 79–104.

Ales, J., Carney, T., & Klein, S. A. (2010). The folding fingerprint of visual cortex reveals the timing of human V1 and V2. *Neuroimage, 49,* 2494–2502.

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *Journal of Neuroscience, 29,* 10638–10652.

Brecelj, J., Kakigi, R., Koyama, S., & Hoshiyama, M. (1998). Visual evoked magnetic responses to central and peripheral stimulation: Simultaneous VEP recordings. *Brain Topography, 10,* 227–237.

Brewer, A. A., Liu, J., Wade, A. R., & Wandell, B. A. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature Neuroscience, 8,* 1102–1109.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research. Brain Research Reviews, 36,* 96–107.

Carlson, T., Hogendoorn, H., Fonteijn, H., & Verstraten, F. A. (2011). Spatial coding and invariance in object-selective cortex. *Cortex, 47,* 14–22.

Chait, M., Poeppel, D., de Cheveigne, A., & Simon, J. Z. (2007). Processing asymmetry of transitions between order and disorder in human auditory cortex. *Journal of Neuroscience, 27,* 5207–5214.

Cichy, R. M., Chen, Y., & Haynes, J. D. (2011). Encoding the identity and location of objects in human LOC. *Neuroimage, 54,* 2297–2307.

Cole, G. G., Kentridge, R. W., Gellatly, A. R., & Heywood, C. A. (2003). Detectability of onsets versus offsets in the change detection paradigm. *Journal of Vision, 3*(1):3, 22–31, http://www.journalofvision.org/content/3/1/3, doi:10.1167/3.1.3. [PubMed] [Article]

de Cheveigne, A., & Simon, J. Z. (2007). Denoising based on time-shift PCA. *Journal of Neuroscience, Methods, 165,* 297–305.

DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., et al. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America, 93,* 2382–2386.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11,* 333–341.

DiCarlo, J. J., & Maunsell, J. H. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of Neurophysiology, 89,* 3264–3278.

Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping, 15,* 95–111.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. A. (2001). Cortical area selective for visual processing of the human body. *Science, 293,* 2470–2473.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (654 pp.). New York: Wiley.

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex, 7,* 181–192.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598–601.

Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science, 330,* 845–851.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics, 36,* 193–202.

Fylan, F., Holliday, I. E., Singh, K. D., Anderson, S. J., & Harding, G. F. (1997). Magnetoencephalographic investigation of human cortical area V1 using color stimuli. *Neuroimage, 6,* 47–57.

Garrido, M. I., Kilner, J. M., Kiebel, S. J., & Friston, K. J. (2007). Evoked brain responses are generated by feedback loops. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 20961–20966.

Gellatly, A., & Cole, G. (2000). Accuracy of target detection in new-object and old-object displays. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 889–899.

Gellatly, A., Cole, G., & Blurton, A. (1999). Do equiluminant object onsets capture visual attention? *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1609–1624.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24,* 187–203.

Hagler, D. J., Jr., Halgren, E., Martinez, A., Huang, M., Hillyard, S. A., & Dale, A. M. (2009). Source estimates for MEG/EEG visual evoked responses constrained by multiple, retinotopically-mapped stimulus locations. *Human Brain Mapping, 30,* 1290–1309.

Hagler, D. J., Jr., & Sereno, M. I. (2006). Spatial maps in frontal and prefrontal cortex. *Neuroimage, 29,* 567–577.

Hemond, C. C., Kanwisher, N. G., & Op de Beeck, H. P. (2007). A preference for contralateral stimuli in human object- and face-selective cortex. *PLoS One, 2,* e574.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36,* 791–804.

Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultra-rapid face detection. *Journal of Vision, 8*(12):9, 1–13, http://www.journalofvision.org/content/8/12/9, doi:10.1167/8.12.9. [PubMed] [Article]

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310,* 863–866.

Im, C. H., Gururajan, A., Zhang, N., Chen, W., & He, B. (2007). Spatial resolution of EEG cortical source imaging revealed by localization of retinotopic organization in human primary visual cortex. *Journal of Neuroscience Methods, 161,* 142–154.

Jeffreys, D. A., & Axford, J. G. (1972). Source locations of pattern-specific components of human visual evoked potentials: I. Component of striate cortical origin. *Experimental Brain Research, 16,* 1–21.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance* (pp. 187–203). Hillsdale, NJ: Lawrence Erlbaum.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics, 43,* 346–354.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

Kastner, S., DeSimone, K., Konen, C. S., Szczepanski, S. M., Weiner, K. S., & Schneider, K. A. (2007). Topographic maps in human frontal cortex revealed in memory-guided saccade and spatial working-memory tasks. *Journal of Neurophysiology, 97,* 3494–3507.

Kouijzer, W. J. J., Stok, C. J., Reits, D., Dunajski, Z., Da Silva, F. H., & Peters, M. J. (1985). Neuromagnetic fields evoked by a patterned on-offset stimulus. *IEEE Transactions on Biomedical Engineering, 32,* 455–458.

Kravitz, D. J., Kriegeskorte, N., & Baker, C. I. (2010). High-level visual object representations are constrained by position. *Cerebral Cortex, 20,* 2916–2925.

Kravitz, D. J., Vinson, L. D., & Baker, C. I. (2008). How position dependent is visual object recognition? *Trends in Cognitive Sciences, 13,* 114–122.

Kriegeskorte, N. (2011). Pattern-information analysis: From stimulus decoding to computational-model testing. *Neuroimage, 56,* 411–421.

Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron, 62,* 281–290.

Liu, L., & Ioannides, A. A. (2006). Spatiotemporal dynamics and connectivity pattern differences between centrally and peripherally presented faces. *Neuroimage, 31,* 1726–1740.

Liu, L., & Ioannides, A. A. (2010). Emotion separation is completed early and it depends on visual field presentation. *PLoS One, 5,* e9790.

MacEvoy, S. P., & Epstein, R. A. (2007). Position selectivity in scene- and object-responsive occipitotemporal regions. *Journal of Neurophysiology, 98,* 2089–2098.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America, 92,* 8135–8139.

Mondy, S., & Coltheart, V. (2000). Detection and identification of change in naturalistic scenes. *Visual Cognition, 7,* 281–296.

Nakamura, A., Kakigi, R., Hoshiyama, M., Koyama, S., Kitamura, Y., & Shimojo, M. (1997). Visual evoked cortical magnetic fields to pattern reversal stimulation. *Cognitive Brain Research, 6,* 9–22.

Niemeier, M., Goltz, H. C., Kuchinad, A., Tweed, D. B., & Vilis, T. A. (2005). Contralateral preference in the lateral occipital area: Sensory and attentional mechanisms. *Cerebral Cortex, 15,* 325–331.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10,* 424–430.

Op De Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology, 426,* 505–518.

Parker, D. M., Salzen, E. A., & Lishman, J. R. (1982). Visual-evoked responses elicited by the onset and offset of sinusoidal gratings: Latency, waveform, and topographic characteristics. *Investigative Ophthalmology & Visual Science, 22,* 675–680.

Phillips, W. A., & Singer, W. (1974). Function and interaction of on and off transients in vision: I. Psychophysics. *Experimental Brain Research, 19,* 493–506.

Poghosyan, V., & Ioannides, A. A. (2007). Precise mapping of early visual responses in space and time. *Neuroimage, 35,* 759–770.

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science, 310,* 1963–1966.

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision, 40,* 49–71.

Portin, K., Vanni, S., Virsu, V., & Hari, R. (1999). Stronger occipital cortical activation to lower than upper visual field stimuli. Neuromagnetic recordings. *Experimental Brain Research, 124,* 287–294.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025.

Robbe, G. L. T., & Op de Beeck, H. P. (2009). Neural representations that support invariant object recognition. *Frontiers in Computational Neuroscience, 3,* 1–16.

Rolls, E. T. (1991). Neural organization of higher visual functions. *Current Opinion in Neurobiology, 1,* 274–278.

Rust, N. C., & Dicarlo, J. J. (2010). Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *Journal of Neuroscience, 30,* 12978–12995.

Sayres, R., & Grill-Spector, K. (2008). Relating retinotopic and object-selective responses in human lateral occipital cortex. *Journal of Neurophysiology, 100,* 249–267.

Schira, M. M., Tyler, C. W., Spehar, B., & Breakspear, M. (2010). Modeling magnification and anisotropy in the primate foveal confluence. *PLoS Computational Biology, 6,* e1000651.

Schwartz, E. L. (1977). Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception. *Biological Cybernetics, 21,* 181–194.

Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 4447–4452.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science, 268,* 889–893.

Sereno, M. I., Pitzalis, S., & Martinez, A. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science, 294,* 1350–1354.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 6424–6429.

Silver, M. A., & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in Cognitive Sciences, 13,* 488–495.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74,* 1–29.

Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *Journal of Neuroscience, 29,* 1565–1572.

Supek, S., Aine, C. J., Ranken, D., Best, E., Flynn, E. R., & Wood, C. C. (1999). Single vs. paired visual stimulation: Superposition of early neuromagnetic responses and retinotopy in extrastriate cortex in humans. *Brain Research, 830,* 43–55.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics, 49,* 83–90.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520–522.

Ungerleider, L. G., & Mishkin, M. (1982). *Two visual pathways* (pp. 549–586). Cambridge, MA: MIT Press.

VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research, 42,* 2593–2615.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron, 56,* 366–383.

Williams, M. A., Baker, C. I., Op de Beeck, H. P., Shim, W. M., Dang, S., Triantafyllou, C., et al. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature Neuroscience, 11,* 1439–1445.