

The Temporal Evolution of Coarse Location Coding of Objects: Evidence for Feedback

Ramakrishna Chakravarthi¹, Thomas A. Carlson², Julie Chaffin³,
Jeremy Turret⁴, and Rufin VanRullen^{3,5}

Abstract

■ Objects occupy space. How does the brain represent the spatial location of objects? Retinotopic early visual cortex has precise location information but can only segment simple objects. On the other hand, higher visual areas can resolve complex objects but only have coarse location information. Thus coarse location of complex objects might be represented by either (a) feedback from higher areas to early retinotopic areas or (b) coarse position encoding in higher areas. We tested these alternatives by presenting various kinds of first- (edge-defined) and second-order (texture) objects. We applied multivariate classifiers to the pattern of EEG amplitudes across the scalp at a range of time points to trace the temporal dynamics of coarse location representation. For edge-defined objects, peak classification performance was high and early and thus

attributable to the retinotopic layout of early visual cortex. For texture objects, it was low and late. Crucially, despite these differences in peak performance and timing, training a classifier on one object and testing it on others revealed that the topography at peak performance was the same for both first- and second-order objects. That is, the same location information, encoded by early visual areas, was available for both edge-defined and texture objects at different time points. These results indicate that locations of complex objects such as textures, although not represented in the bottom-up sweep, are encoded later by neural patterns resembling the bottom-up ones. We conclude that feedback mechanisms play an important role in coarse location representation of complex objects. ■

INTRODUCTION

Determining the location of a given object is one of the central tasks of the visual system. The classic view is that the visual system employs a division of labor strategy, with the ventral stream taking on the task of identifying objects (the “what” pathway) and the dorsal stream involved in determining their locations for enabling action (the “where” pathway; Milner & Goodale, 1993; Ungerleider & Mishkin, 1982). However, recent studies have shown that this dichotomy is not airtight, with position and motion also being processed by the ventral stream (Li et al., 2013; DiCarlo & Maunsell, 2003); in fact, the dichotomy might not hold up upon closer scrutiny (Hesse, Ball, & Schenk, 2012; Schenk & McIntosh, 2009). Beyond the debate over where object locations might be encoded, two crucial questions remain unanswered: (1) What is the nature of location representation; specifically, are the locations of all objects represented in the same way or do they depend on where their identity might be processed in the ventral stream? (2) How do these representations evolve in the visual system over time? The current study is aimed at addressing these questions by applying multivariate

pattern analysis techniques to EEG traces obtained when participants viewed different kinds of objects.

To address the question of the nature of location representation, we utilized six different kinds of objects. These belonged to two categories: first-order (or edge-defined) objects and second-order (or texture-defined) objects. It is known that these objects are processed in different areas of the ventral stream. For example, orientation and contrast defined second-order objects require processing in higher visual areas, such as V4 and LOC, for segmentation, whereas first-order objects can be processed as early as in V1 (Thielscher, Kollé, Neumann, Spitzer, & Gron, 2008; Larsson, Landy, & Heeger, 2006; Kastner, De Weerd, & Ungerleider, 2000). Furthermore, within each category, we used objects that require differentiable processing mechanisms, most likely occurring in different areas. The first-order objects we chose are processed at different speeds. We used objects with high luminance, low luminance, and isoluminant colors. It is known that luminance (and hence contrast) affects the neuronal latencies of even early visual areas (Gawne, Kjaer, & Richmond, 1996; Lennie, 1981). Furthermore, luminance defined first-order objects stimulate both parvo- and the faster magnocellular pathways, whereas isoluminant first-order color objects would only stimulate the slower parvocellular mechanisms (Schiller & Logothetis, 1990; Livingstone &

¹University of Aberdeen, ²Macquarie University, Sydney, New South Wales, Australia, ³Université de Toulouse, ⁴University of Maryland, ⁵Centre National de la Recherche Scientifique, Toulouse, France

Hubel, 1988). Similarly, we chose second-order textures that would require different amounts of additional processing steps. We tested three types of textures: orientation-defined, identity-defined pop-out, and identity-defined conjunction textures. Orientation is processed earlier in the hierarchy than identity. Furthermore, textures that pop out are preattentively processed, whereas those that do not pop out require attention for object identification (Treisman & Gelade, 1980). Thus, the second-order objects require processing in different visual areas with the conjunction stimulus requiring the furthest processing. We used a variegated set of objects to determine if location representation would be uniform across the entire set or if their particular ventral stream processing requirements would determine their location representation.

Where might the location of such objects be represented? Early visual areas have a well-defined retinotopic layout (Felleman & Van Essen, 1991; Essen & Zeki, 1978). This map-like organization can be the obvious place for the visual system to encode object locations. However, this might not be possible for all objects. First, not all objects elicit unambiguous, spatially localized activity in the early visual cortex. For example, texture-defined objects do not activate striate cortex in a localized manner but are processed only in later areas. Second, there is an ongoing debate about the appropriate coordinate system—retinotopic or spatiotopic—for visual perception (Golomb & Kanwisher, 2012; Cavanagh, Hunt, Afraz, & Rolfs, 2010; Gardner, Merriam, Movshon, & Heeger, 2008; d'Avossa et al., 2007; Melcher & Morrone, 2003). The retinotopic layout of early visual cortex might not be sufficient if the visual system relies on spatiotopic coordinates. Finally, higher visual areas involved in object recognition are either position invariant (Rust & Dicarlo, 2010; Ito, Tamura, Fujita, & Tanaka, 1995; Logothetis & Pauls, 1995; Schwartz, Desimone, Albright, & Gross, 1983) or encode position information only in a very coarse manner (Carlson, Hogendoorn, Fonteijn, & Verstraten, 2011; Cichy, Chen, & Haynes, 2011; Macevoy & Epstein, 2007; DiCarlo & Maunsell, 2003). These findings suggest that later ventral areas, which are needed for object identification, have, at best, some location information. Thus, early visual cortex can encode precise location information but cannot isolate complex objects, whereas higher visual areas can isolate complex objects but can only represent their coarse location. Tasks that require coarse localization of objects can therefore utilize either the location information present in these higher areas or the precise location information available in the retinotopic cortex. Here, we will determine if differences in object processing translate to differences in the representation of their coarse location. Two possibilities exist for encoding coarse location of objects in the ventral stream: the feed-forward and the feedback hypotheses.

According to the feed-forward hypothesis, an object's representation, including its location, is computed on the fly as various areas are activated in the feed-forward sweep.

If so, location representation will differ for different objects, depending on where in the visual hierarchy it is isolated. There is ample evidence that visual stimuli are processed in a hierarchical manner for object recognition. For example, studies on ultrarapid categorization (Fei-Fei, Iyer, Koch, & Perona, 2007; Kirchner & Thorpe, 2006; Thorpe, Fize, & Marlot, 1996) have shown that objects can be categorized using purely feed-forward processing. Intracranial recordings have demonstrated that objects can be categorized and identified on the basis of just the earliest spikes elicited by the stimulus, suggesting that feed-forward processing is sufficient for detailed object processing. Accurate classification of object category can be performed as early as 100 msec after stimulus onset in monkeys (Hung, Kreiman, Poggio, & DiCarlo, 2005) and humans (Liu, Agam, Madsen, & Kreiman, 2009). These and similar findings have inspired successful models of the visual system that emphasize feed-forward hierarchical processing (see, e.g., Serre & Poggio, 2010; Serre et al., 2007; Ullman, 2007; Oliva & Torralba, 2006; VanRullen & Thorpe, 2002; Edelman & Intrator, 2000; Wallis & Rolls, 1997). One such influential model, developed by DiCarlo, Zoccolan, and Rust (2012), proposes a purely feed-forward mechanism for “core” object recognition. According to this model, patterns of activity in a population of IT neurons receiving inputs from early visual areas can identify objects in an invariant manner—that is, irrespective of their position, size, viewpoint, etc. However, individual IT neurons that contribute to this population coding can be highly selective to a given object (or a complex combination of features) and retain its position (and size, viewpoint, etc.) information (Zoccolan, Kouh, Poggio, & DiCarlo, 2007). That is, a given IT neuron responds to a particular object in a (roughly) specified position. DiCarlo et al. (2012) posit that the position information available in these neurons can then be utilized by downstream circuits for coarse localization of objects (“was this object in the left visual field?”) The population of neurons therefore can invariantly identify an object without throwing out other relevant information. As DiCarlo et al. (2012) put it, “The resulting [IT] population representation is powerful because it simultaneously conveys explicit information about object identity and its particular position, size, pose, and context.” In short, as each region successively computes a more complex representation, it can also simultaneously encode the object's coarse location. This is further supported by the findings that downstream neurons (even those with large receptive fields) are capable of tracking coarse location (Aggelopoulos & Rolls, 2005; DiCarlo & Maunsell, 2003; Op De Beeck & Vogels, 2000; Schwartz et al., 1983) and that saccades can be directed very quickly (within 100–200 msec) to the location of the stimulus belonging to the appropriate stimulus category (Drewes, Trommershauser, & Gegenfurtner, 2011; Crouzet, Kirchner, & Thorpe, 2010; Kirchner & Thorpe, 2006).

Alternatively, the feedback hypothesis suggests that higher visual areas initiate feedback loops to early visual

areas enabling them to keep track of the object's location. According to this hypothesis, therefore, the representation of position will be similar across all objects, if they occupy the same visual location, irrespective of their identity. However, depending on the complexity of the object and the depth of processing needed to determine its identity, the position information might be accessed from early visual areas at different times. There is abundant evidence that early visual cortex is actively involved in processing complex stimuli. For example, Zipser, Lamme, and Schiller (1996) demonstrated that V1 neurons fire differently depending on whether the stimulus patch within their receptive fields belongs to a figure or the ground, even when low-level stimulus properties are kept identical. This modulation appears to occur at least 100 msec after target onset, suggesting that feedback processes contribute to this modulation. Further experiments (Roelfsema, Lamme, & Spekreijse, 2004; Lamme, Zipser, & Spekreijse, 2002; Hupe et al., 1998, 2001; Pascual-Leone & Walsh, 2001; Super, Spekreijse, & Lamme, 2001a, 2001b; Lamme, Rodriguez-Rodriguez, & Spekreijse, 1999; Rao & Ballard, 1999) suggested that feedback to early visual cortex was involved in several visual processes such as texture segregation, visual STM, backward masking, grouping, among others. However, it could be that these feedback modulations were only by-products, reflecting increased activity in higher cortical areas. Recent studies using TMS, a technique that temporarily disrupts neuronal activity, have demonstrated that feedback to early visual cortex (V1, specifically) is causally implicated in many visual processes (Wokke, Sligte, Steven Scholte, & Lamme, 2012; Heinen, Jolij, & Lamme, 2005; Juan & Walsh, 2003; Ro, Breitmeyer, Burton, Singhal, & Lane, 2003). In summary, feedback projections to early visual cortex seem to play a significant role in processing objects. Although coarse location can be coded in higher visual areas by feed-forward processing, it is plausible that location coding is still dependent on accessing the detailed information available in early visual cortex through feedback (see, e.g., Bullier, 2001).

In this study, we tested these alternatives—the feed-forward and feedback hypotheses—by applying classification techniques to EEG (see Carlson, Tovar, Alink, & Kriegeskorte, 2013; Carlson, Hogendoorn, Kanai, Mesik, & Turret, 2011) while participants viewed various types of objects. This allows us to extract information about object location from ongoing neural activity on a millisecond-by-millisecond basis, thus tracing the time course of location representation. According to both feed-forward and feedback hypotheses, locations should be represented at different time points for the different types of objects, depending on where in the ventral stream they are processed. However, if the feed-forward hypothesis is correct, the information content of these various representations should be different from each other. On the other hand, the feedback hypothesis predicts that the information content for all representations should be the same. This is because, although objects are processed

in different visual areas, their location is specified in the earlier retinotopic areas.

To test these hypotheses, we adopted a simple task to be performed on the various object types. We presented one object at a time either to the left or to the right of fixation and applied multivariate pattern analysis classifiers on the resulting EEG to determine if the coarse location of the object (left or right visual field) can be predicted. We then looked at the time course of these predictions. We also tested if the location information for each object was the same or different. To ensure that the participants paid attention to the stimulus, we engaged them with an object identification task (was the object a circle or a square). This task had the additional virtues that it (a) was orthogonal to the main question of the study (left or right) and hence any signals arising from this task, particularly motor activity, would not contaminate the classification procedure and (b) ensures that the identification process occurs, which is thought to be mediated by the ventral stream, since we want to know if this process influences location encoding.

METHODS

Observers and Stimuli

Eleven volunteers (aged 21–37 years) with normal or corrected-to-normal vision participated in this experiment with informed consent. Objects were generated using MATLAB with the Psychtoolbox extensions (Kleiner, Brainard, & Pelli, 2007; Brainard, 1997; Pelli, 1997) and presented on a cathode ray monitor with a resolution of 800×600 pixels and a frame rate of 100 Hz. The display was placed 57 cm from the observer, whose head was stabilized with a chin and forehead rest.

The set of stimuli consisted of three first-order and three second-order objects (Figure 1B). Each object had two shape versions—a square and a circle of size 5 deg, width and diameter, respectively. The first-order objects had a different luminance or color relative to a uniform gray background (luminance = 11.7 cd/m^2). The three first-order stimuli were characterized by (a) high luminance (36.2 cd/m^2 ; Michelson contrast = 51%), (b) low luminance (17.3 cd/m^2 ; contrast = 12%), and (c) isoluminant color (green), respectively. Each first-order stimulus had a clearly discernable edge in contrast to the second-order objects. The latter were defined by orientation or identity differences of object elements relative to background elements. The three second-order objects were textures characterized by (a) orientation: horizontal lines embedded among vertical lines or vice versa, (b) pop-out: O's embedded among + 's or vice versa, and (c) conjunction: L's embedded among T's or vice versa, respectively. The lines in the orientation texture were 1.4×0.1 deg and the letters in the other two textures were 0.7×0.7 deg in size, and all had high luminance (36.2 cd/m^2). The lines were placed in random locations

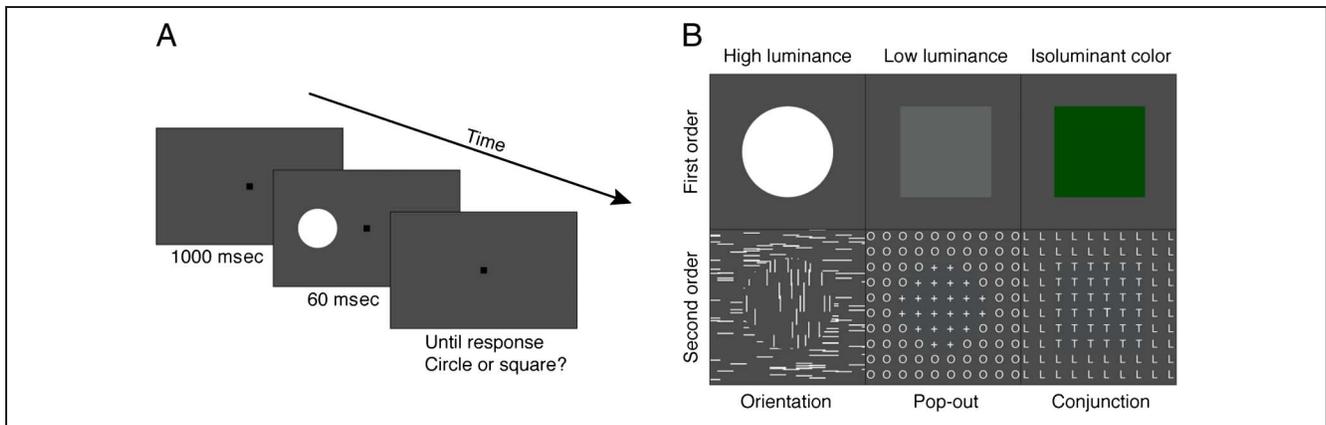


Figure 1. Procedure and objects. (A) Timeline of a single trial. An object is presented briefly (60 msec), and the observer is asked to report its shape. (B) Examples of the objects that were used in the experiment. The top row shows the various first-order (edge-defined) objects, and the bottom row shows the second-order (texture-defined) objects. The background elements for texture objects covered the entire screen.

with the restriction that they do not overlap each other; the letters were spaced 1.1 deg apart in a regular grid. On each trial, the identity of the object element (orientation of the lines or the identity of the letters) was picked randomly. The background elements had the complementary identity and covered the entire screen such that there was no discernible edge. In orientation and pop-out textures, the identities of the object and background elements differed by a single feature, whereas in the conjunction stimulus, the features of object and background elements were shared but differed in their conjunction.

Experimental Procedure

First, we determined the isoluminant green for each observer using heterochromatic flicker photometry (Ives & Kingsbury, 1914). A green object at a randomly selected luminance was presented at an eccentricity of 10 deg to the left or right of a black fixation square (size 0.7 deg) along the horizontal meridian. The object was created by activating only the green channel of the monitor, with the other two color channels providing no output. The object was either a square or a circle of size 5 deg (width or diameter). It flickered at 15 Hz with each cycle consisting of two frames: one frame contained the object embedded in the gray background and the other only the background (square-wave modulation). The observer was asked to adjust the luminance of the flickering object using key presses, while maintaining fixation on the fixation square, until the perception of flicker was minimal or absent. Ten trials were run with five trials for each location (left/right). The average of the reported luminance across the five trials for each location was taken as the observer's isoluminant point for green for that location.

In the main experiment, we tested six object types, described above. There were 210 trials per object type, with 105 trials on each side of fixation, for a total of 1260 trials per observer. The order of object type and location was

randomized. The shape of the object was randomly chosen on each trial. Each trial was as follows (Figure 1A). A key press initiated the trial; 1 sec later, an object was presented for 60 msec at an eccentricity of 10 deg along the horizontal meridian. The observer was asked to maintain fixation during this period and report the shape of the object as quickly as possible with a key press. The response and the RT were recorded. Visual feedback was provided on each trial. EEG was simultaneously recorded. Observers were asked to perform a task orthogonal to the question of interest: How is object position encoded? They performed a shape discrimination task and did not report object location to ensure that (a) motor signals indicating left or right locations did not inadvertently contaminate the data, (b) they were paying attention to the stimulus, and (c) they engaged in an identification task, which is a key variable in testing the two hypotheses. We chose to present the objects to the left or right of fixation in different hemifields because this requires only coarse localization of the objects. Higher visual areas can reliably encode the hemifield locations of objects that they isolate (Carlson, Hogendoorn, Fonteijn et al., 2011; Cichy et al., 2011; Hemond, Kanwisher, & Op de Beeck, 2007; Macevoy & Epstein, 2007; DiCarlo & Maunsell, 2003). This allowed us to directly test the two hypotheses regarding the nature and time course of coarse location representation in the brain.

EEG Acquisition

Sixty-four-channel EEG was recorded using a BioSemi Active Two system at a sampling rate of 1024 Hz, which was later downsampled to 256 Hz. A three-channel EOG was also recorded to monitor eye movements and blinks. Epochs were created ranging from 300 msec before to 600 msec after stimulus onset. Baseline correction was applied by subtracting the average activity, from each channel, between -300 and 0 msec relative to stimulus onset. EEG data were screened manually on a trial-by-trial basis,

rejecting all trials with visible artifacts, eye movements, and blinks.

Trial-by-Trial Pattern Classification

The goal of classification was to determine if there was sufficient information in the scalp distribution of EEG amplitudes, on a moment-to-moment basis, to predict the location of the presented object. This analysis probes the temporal evolution of coarse location representation for that object. If a classifier successfully predicts the location of the object at a given time point, then location information is available to the visual system at that time point.

We constructed a classifier for each time point and object. Each support vector machine (SVM) classifier utilized a nonlinear radial basis function (RBF) kernel implemented in the LIBSVM library (www.csie.ntu.edu.tw/~cjlin/libsvm; Chang & Lin, 2011) with default c (cost, set to 1) and γ (set to the reciprocal of the number of features, here $1/64$) parameter values. For each object, 90% of trials were assigned to a training set. One SVM classifier per time point was trained, with the appropriate location labels, on z -scored (scaled) 64-channel EEG amplitudes obtained at that time point of these trials. The trained classifier then predicted the location of the remaining 10% of trials, the test set, on a trial-by-trial basis using the scaled EEG amplitudes obtained on these trials at the corresponding time point. A 10-fold cross-validation was conducted by repeating this procedure on nonoverlapping sets of test trials and the corresponding training trials, thus ensuring that each trial was tested once. The classifier's predictive performance was averaged across all cross-validation runs to obtain the time course of location prediction for a given object type. We also determined the weights attributed to each electrode by the classifiers and therefore the time course of the contribution of each electrode to the classifiers' performance. We computed baseline performance by applying the same procedure to the same EEG data, but with location labels shuffled. If any systematic bias was inherent to the classification algorithm, then it should be manifest in the baseline computation.

ERP Analysis

For each object type, location, and electrode, we obtained ERPs by averaging the EEG activity time-locked to stimulus onset. Then for each object type and location, we computed the "contralateral – ipsilateral" ERP difference at occipital electrodes (O1 and O2). For each object, we obtained differential ERPs by averaging the above difference across the two target locations.

Other Classification Algorithms

In this study, we used an SVM pattern classification algorithm with a nonlinear RBF kernel with default parameters.

To test if our results were specific to this algorithm, we also analyzed the data with (a) linear discriminant analysis algorithm, (b) linear discriminant analysis preceded by a principal component analysis applied to the EEG data to reduce the feature set, and (c) a nonlinear RBF kernel in the SVM algorithm with optimized cost and gamma parameters determined by a thorough search of the parameter space for each observer. The results from these classification methods were indistinguishable from our default analysis, with these classifiers even performing slightly better at some time points. We present, here, only the data from our analysis with SVM classifiers using an RBF kernel with default parameters.

Source Localization

The two hypotheses predict different sites for coarse location representation, with the feed-forward hypothesis suggesting successive areas along the ventral pathway and the feedback hypothesis implicating early visual areas. To test this, we estimated the sources of relevant scalp activity. We merged the epoched EEG data of all participants into one large data set. We then divided this combined EEG data set into 12 smaller data sets according to object type (6) and location (2) combinations. The following procedure was run for each of those data sets such that a source was estimated for each object type and location. Independent component analysis was run on the data set, and the components were sorted by the amount of variance they explained. To determine the sources of activity relevant for location representation, the time point at which the classifier for that object type performed best was noted. We then isolated EEG activity 50 msec on either side of this peak and obtained 100-msec strips of EEG activity. This EEG activity contains that object's location representation. The independent component that best explained this ERP activity and was also among the top 10 components (in terms of variance explained) was identified. It is reasonable to consider that this component underlies the location representation. A dipole was then fit to this component using the "DIPFIT2" function in the EEGLAB toolbox (Delorme & Makeig, 2004) with default parameters and a spherical four-shell BESA head model. The estimated dipole, with its location and moment, was plotted in Talairach coordinates. Using source localization based on independent component analysis allows us to isolate the sources of location-specific neural activity in an independent manner (instead of relying on grand-averaged waveforms that might potentially reflect many other concurrent processes).

RESULTS

Behavioral Performance

Table 1 summarizes the behavioral outcomes for human observers when asked to identify the shapes of various

Table 1. Classification and Behavioral Performance

Stimulus Type	Human Performance (Shape Discrimination)				Classifier Performance (Localization)				
	RT (msec)	SEM	d'	SEM	Onset Latency (msec)	Peak Latency (msec)	SEM	Peak Performance (% Correct)	SEM
High luminance	583	27	4.04	0.25	59	131	4	80	2.2
Low luminance	600	27	4.18	0.26	66	141	3	80	2.2
Isoluminant color	614	25	3.96	0.21	94	184	14	83	1.7
Mean first-order	598	26	4.06	0.21	73	152	8	81	2
Orientation	755	29	2.07	0.13	160	263	25	68	1.7
Pop-out	831	32	0.66	0.09	196	307	24	69	2
Conjunction	855	46	0.32	0.1	223	314	19	64	1.4
Mean second-order	813	39	1.02	0.09	193	295	23	67	1.7

This table lists human and classifier response measures on orthogonal tasks (shape discrimination and location prediction, respectively) performed on the same stimulus set.

objects. All reported p values have been Bonferroni-corrected for multiple comparisons as and when needed. Observers identified first-order objects faster (mean RT = 598 ± 26 msec; paired t test: $t(10) = 102.6, p < 10^{-7}$) and more accurately (mean $d = 4.06 \pm 0.21$; paired t test: $t(10) = 213.5, p < 10^{-7}$) than second-order objects (mean RT = 813 ± 39 ; mean $d = 1.02 \pm 0.09$). Responses to the three first-order objects were equally accurate (one-way repeated-measures ANOVA: $F(2, 20) = 0.6, p > .5$) but had different RTs, $F(2, 20) = 13.4, p = .0002$. However, responses to second-order objects differed both in speed ($F(2, 20) = 12.6, p = .0003$) and accuracy ($F(2, 20) = 210.2, p < 10^{-7}$). Post hoc comparisons (please see Table 1 for mean RTs and d s for each object type) showed that observers were faster and more accurate for orientation textures than for either pop-out (RT: $p = .026$; d : $p < .0001$) or conjunction textures (RT: $p = .007$; d : $p < .0001$). Similarly, they were marginally faster ($p = .059$) and more accurate ($p = .008$) for pop-out than for conjunction textures. In summary, observers could rapidly and accurately determine the identity of objects that had a clearly defined edge, but they took longer and made more errors when presented with objects that required further processing. Performance deteriorated with more complex objects.

Classification Performance

Pattern classification on EEG data extracts the time course of coarse location information for each object type (Figure 2A). Each time course can be characterized by its peak performance (highest classifier performance), peak latency (time point at which peak performance is reached), and onset latency (time point when performance first deviates from baseline performance, as determined by paired

t tests; see Figure 2). Overall, as noted in Table 1, classifier peak performance for the first-order objects was high, and the corresponding peak latencies were short. On the other hand, prediction was weaker and later for second-order objects (paired t test, peak performance: $t(10) = 61.5, p < .0001$; peak latencies: $t(10) = 97.8, p < .0001$). Among first-order objects, post hoc comparisons showed that peak latencies for high- and low-luminance objects were the same ($t(10) = -0.73, p > .5$), suggesting that luminance does not play a major role in location representation, whereas peak latencies for isoluminant color objects were delayed (vs. high luminance: $t(10) = -2.88, p = .049$; vs. low luminance: $t(10) = -2.71, p = .065$). This is consistent with the fact that isoluminant color objects activate only the parvocellular pathways whereas luminance objects also activate magnocellular pathways (Schiller & Logothetis, 1990; Livingstone & Hubel, 1988). It is known that, because of this differential activation, neuronal latencies for chromatic stimuli are slower than for achromatic stimuli by around 20–40 msec (Maunsell et al., 1999; Nowak, Munk, Girard, & Bullier, 1995; Maunsell, 1987), similar to what we found in this study.

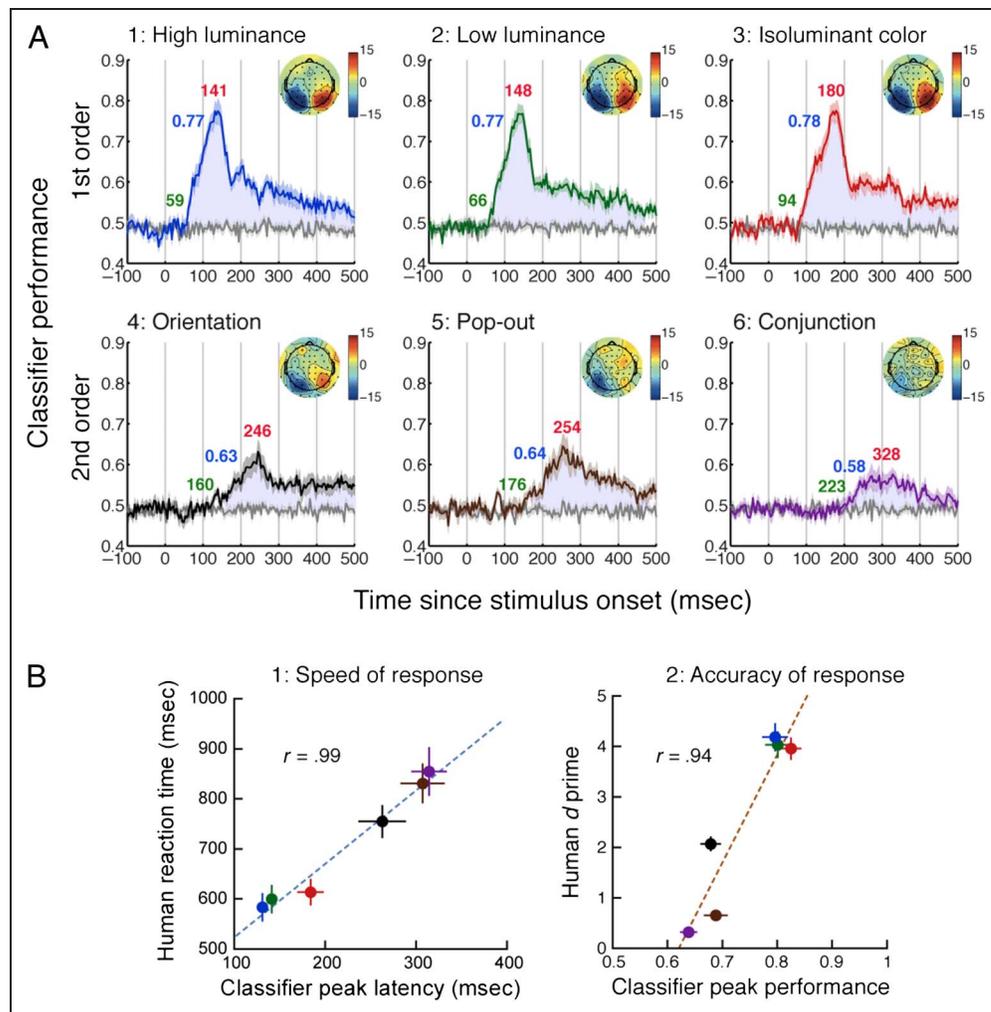
Analogous to the behavioral results, there were no significant differences in the classifiers' peak performance for first-order objects ($F(2, 20) = 1.88, p = .18$). However, peak performances were different among the second-order objects ($F(2, 20) = 5.22, p = .015$). Post hoc comparisons showed that there was a performance drop for conjunction textures compared with orientation textures ($p = .03$) and pop-out textures ($p = .09$). These results suggest that the more processing a second-order stimulus needs, the lower the classifier's performance. These time-course findings are compatible with both feed-forward and feedback hypotheses of location representation. Second-order objects have to be processed further than first-order objects for segregation. Therefore, location is represented later

because of corresponding (a) increases in processing times or (b) delays in feedback.

Although the observers were asked to perform an orthogonal task (shape discrimination), their mean accuracy (d_s) and mean RTs were highly correlated with the classifiers' mean peak performance ($r = .94, p = .0045$) and mean peak latency ($r = .99, p = .0043$), respectively (Figure 2B). It must be noted that this correlation is driven by object type. That is, both observers and classifiers perform similarly for a given object type. This suggests that the observers drew on the same representation to perform the shape discrimination task as the classifier did to predict the object's location. In other words, the representation of coarse location appears to be tightly

coupled to processes that underlie object recognition, which presumably occurs in the ventral visual stream. This is another instance where a putatively independent function—position encoding, classically thought to take place in the dorsal stream—has been found to be closely linked to the processes in the ventral stream, casting doubt on the clean division of labor theory of two visual streams (Schenk & McIntosh, 2009). One interpretation is that the visual system develops a single representation that includes both identity and (coarse) location information. This representation can then be used for various actions and tasks. A corollary to this interpretation is that the pattern of activity utilized by the classifier to successfully predict location is not merely epiphenomenal but

Figure 2. Classification and behavioral performance. (A) Classification performance as a function of time for each of the six object types tested in the experiment. The colored line in each plot is mean classifier performance across participants for that object type. Baseline performance is plotted in gray. Shaded regions represent 1 SEM. Baseline performance is obtained by running the same classifier algorithm but with location labels (left or right) shuffled. The area shaded in light blue within each plot indicates the time points where classifier performance is significantly above baseline. To determine these points, we performed pairwise comparisons at each time point. Any time point with a significant outcome ($p < .05$) that was flanked or followed by at least four consecutive time points that also had significant outcomes (that is, when above chance performance was observed for at least 20 msec) was shaded blue. Red numbers within each plot indicate peak latency—time since stimulus onset at which classifier performance is the highest; green numbers indicate onset latency—the earliest time point at which classifier performance differs from baseline; blue numbers represent the classifier's peak performance values. These values are derived from the time courses averaged across subjects, shown above. The means of subject-wise performance are listed in Table 1. Head plots for each object show the distribution of weights learnt by the classifier at that object's peak latency. The electrode contributions are similar across different objects and are mostly from parieto-occipital activity. (B) Classifier performance against human behavioral performance. Each filled circle represents average performance across participants for one object type. The color of the circle indicates the object type (blue = high luminance, green = low luminance, red = isoluminant color, black = orientation texture, brown = pop-out texture, and purple = conjunction texture). B.1 shows that classifier peak latency and human RTs are highly correlated. Similarly, B.2 shows that classifier peak performance and human accuracy are highly correlated. These findings suggest that both the classifier and the observers draw from the same neural representation to perform their respective tasks.



plays a key role in the visual system's overall processing of objects and can guide behavior.

Testing the Two Hypotheses

In this study, we were interested in the neural mechanisms that underlie coarse location representation. To directly test the two hypotheses—feedback and feed-forward—we looked at whether location information was the same or different across different objects. If location is represented in early visual areas, as the feedback hypothesis posits, then location representation will be similar across object types, despite differences in timing. On the other hand, if location is represented in the feed-forward sweep, the timing and nature of location information will differ according to where each object is isolated in the visual processing stream.

Qualitatively, the current evidence favors the feedback hypothesis, as the electrodes most preferred by each object's classifier tend to be the same ones—the parieto-occipital electrodes (as seen in the head plots in Figure 2A). To quantify this observation and further test the two hypotheses, we trained a classifier on one object and tested it on other objects. If it successfully predicts the locations of untrained object types, then the information content is the same across all objects; otherwise location is represented differently for different objects. In other words, successful prediction implies that the preferred electrodes and the topographical distribution of activity are the same for classifying locations among all object types.

Specifically, we trained a classifier on high-luminance object trials at its peak latency (~140 msec after stimulus onset), the high-luminance classifier, and tested it at all time points of the six object types. This analysis revealed (Figure 3A) that the location of untrained object types could be predicted as well as when the classifier was trained and tested on the same object types themselves, implying that the information available to each classifier was the same across objects. That is, the locations of all tested objects are represented in the same way. In particular, the peak performances and peak latencies of the high-luminance classifier were the same as those obtained in the earlier analysis, where classifiers were trained and tested on the same objects. Crucially, this implies that the information learned by the high-luminance classifier (at 140 msec) becomes available only at later time points for texture objects—the same location information is observed at different time points for each object.

Further evidence of feedback comes from the observation that the high-luminance classifier (solid colored line in Figure 3A) performs as well as the original classifier (dashed colored line), not only at the peak latencies, but also at several later time points, for most object types. Figure 3A shows that performance of the high-luminance classifier falls down to chance shortly after peak perfor-

mance is achieved but then rises to the level of the original classifier at several subsequent time points. That is, the information present in the classifiers trained on second-order objects trials is identical to that present in the high-luminance classifier at multiple time points. This finding suggests that, for a given object, the same location information present at an early time point reappears at later time points, implying feedback of information. Surprisingly, this seems to be the case even for first-order objects, suggesting that feedback processes might be involved even for first-order objects, contrary to the feed-forward hypothesis of object representation.

It is interesting to note that the high-luminance classifier fails to predict the location of objects or even consistently mispredicts the location of objects at certain time points, noticeably right before and after the peak latency; at these time points, the classifier trained and tested on the same object is itself only moderately successful. The successful prediction of the latter probably derives from information present either in parieto-occipital electrodes, but in a form different from the one on which the high-luminance classifier was trained, such as parieto-occipital activity with a different polarity, or from information present in other electrodes not utilized by the high-luminance classifier.

Classification as Scalp Activity Differences

As described earlier, a different way of looking at this data is that the classifiers have learnt the topographic distribution of electrical activity as a template for prediction. Therefore, if the high-luminance classifier successfully predicts the location of untrained object types (e.g., texture-defined objects), then the topography must be the same for high-luminance objects and other objects. This is attested by two related findings. First, parieto-occipital activity was more strongly weighted by all classifiers, irrespective of object type, at their peak latencies (Figure 2A), suggesting that these regions carried the most relevant location information on the scalp. Furthermore, this lends support to the suggestion that location is represented in early visual cortex. Second, differential occipital ERPs (contralateral minus ipsilateral ERPs) closely followed and were strongly correlated with the performance of the high-luminance classifier (Figure 3A). Differential occipital ERPs provide a summary of the occipital topography at each time point. To the extent that occipital activity serves as a template for the high-luminance classifier and an indicator of location information, the classifier's performance will correlate with the differential ERP, irrespective of object identity, which is what we find.

Source Localization

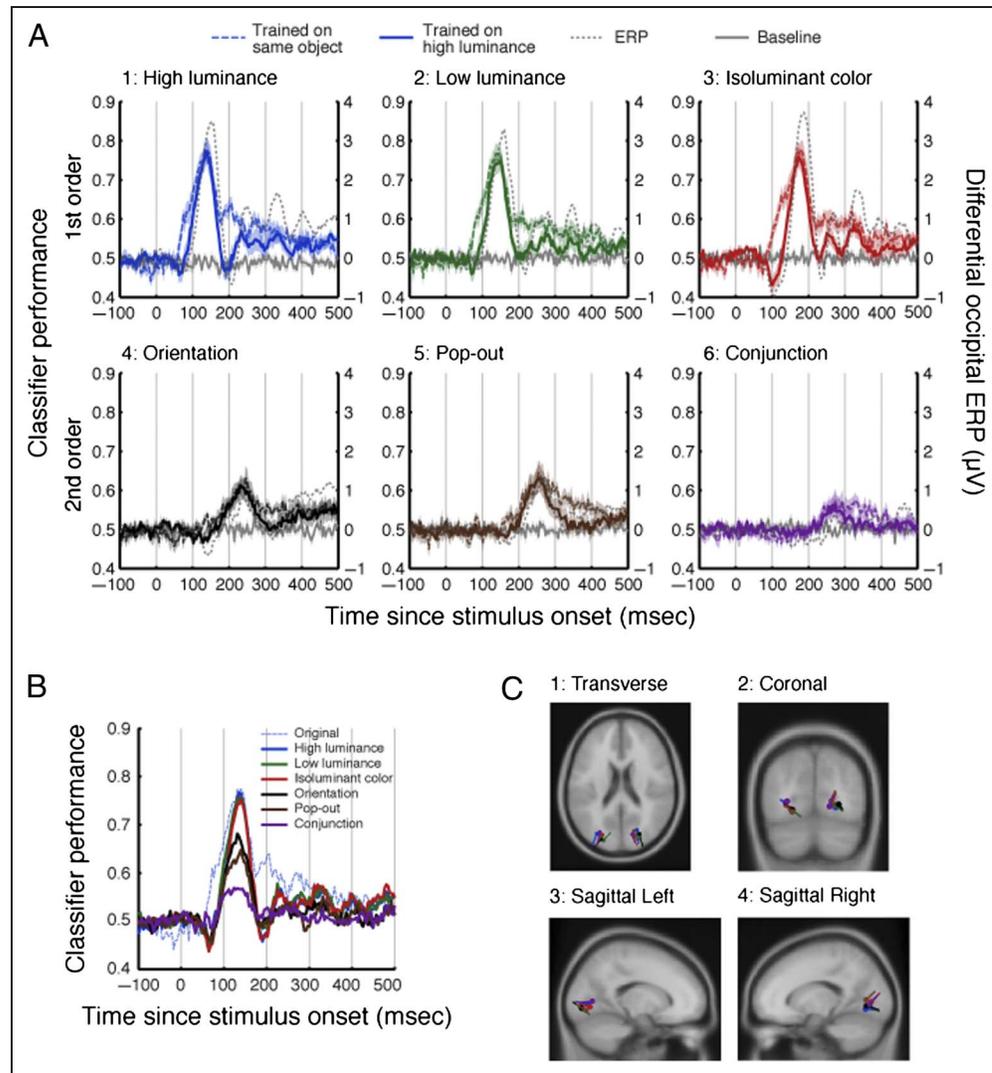
As a direct test of the two hypotheses, we determined the source (position and moment of the best fitting dipole) of the scalp activity relevant to each object's location

Figure 3. Cross-training analyses. To test if coarse location is represented similarly across various objects, we trained a classifier on high-luminance trials at its peak latency—the high-luminance classifier—and tested it at all time points of all objects (A). The solid color lines represent the high-luminance classifier’s performance; the dashed color lines are the original performance traces presented in Figure 2. The high-luminance classifier performs as well as the original classifiers, implying that location representation is shared across object types. Also plotted are differential ERPs (dashed gray lines) for each stimulus: mean difference between contralateral and ipsilateral occipital (O1 and O2) ERPs. These differential ERPs summarize the topography of occipital activity. As seen in the plots, they closely follow the time course of the high-luminance classifier’s performance. (B) A further test of the hypothesis that location is represented in the same way across objects. We trained classifiers at the peak latencies of all objects using the respective object’s trials and tested it on high-luminance trials with each of these classifiers. The resulting plot shows that we can recover the same time course, but with

varying degrees of accuracy. In general, training the classifier on first-order objects allows excellent classifier performance whereas training on second-order objects leads to weak performance (although at the right time points). This suggests that location representation of second-order objects (although similar in topography to that of first-order objects) has a lower signal-to-noise ratio relative to first-order objects. (C) Estimated locations and moments of dipoles fitted to relevant scalp activity. The sources for all object types seem to be concentrated in early visual areas, despite the temporal differences in the activity relevant for their location representation, strongly supporting the feedback hypothesis.

representation—the activity that drove their classifier’s performance. The feed-forward hypothesis predicts that the sources would be distinct and spread out along the ventral visual pathway. On the other hand, if the feedback hypothesis is correct, the sources should all be located in retinotopic or early visual areas. We found that the sources for all object types overlapped considerably, were within a centimeter of each other, and were situated in Brodmann’s areas 17 and 18 (Figure 3C), as predicted by the feedback hypothesis. The DIPFIT2 function implemented in EEGLAB allows coarse scanning and then fine-grained fitting of the dipoles to independent components to estimate sources. We found that this process allowed precise dipole fits; that is, the estimated dipoles explained a large portion of the variance in the selected components whose sources were being estimated. On

average, the dipole fits for the six left sources had a residual (unexplained) variance of $2.8 \pm 0.7\%$ (mean \pm SEM; range = 1.6–5.9%). Similarly, the residual variance for the right dipole fits was $3.1 \pm 0.6\%$ (range = 0.8–5%). Note that the relevant scalp activity, for which the sources were estimated, was culled from different time points for the various object types. Thus, the relevant scalp activity for first-order objects was around 140 ± 50 msec after stimulus onset, whereas that for second-order objects was at least 100 msec after that. Despite these differences in time, the sources were all estimated to be in the same early visual areas, strongly supporting the feedback hypothesis. Although source localization is known to suffer from issues with inverse solutions, these findings are complementary to the findings that classifiers relied heavily on parieto-occipital activity and that



occipital differential ERPs predicted object location and further support the feedback hypothesis.

To further confirm the feedback hypothesis, we performed the reciprocal analysis to the earlier cross-object classification analysis: We trained separate classifiers at the peak latencies of each object type and tested each classifier at all time points of the high-luminance trials (Figure 3B). We found that these classifiers traced the same time course as when the classifier was trained and tested on the high-luminance object. Just as with the source localization results, although the classifiers were trained at different “optimal” time points for different objects, the obtained peak latencies of all classifiers when tested on high-luminance trials were indistinguishable. This again suggests that the same location information is present at the peaks of all objects. However, the various classifiers had different peak performance values, with the classifiers trained on second-order objects having the weakest performance. The reduced performance indicates that the information recovered from second-order objects must have a lower signal-to-noise ratio (SNR) than that of the high-luminance objects and hence constituted a suboptimal template for identifying the location of an object. The lower occipital ERP amplitudes for second-order objects also support the claim that texture objects are represented in the early visual areas with a lower SNR.

In short, the information or topography present at the peak of each stimulus is identical or highly similar, albeit occurring with different SNR and at different time points. Second-order objects cannot stimulate localized regions during the feed forward sweep, as confirmed by the late onset latencies of the differential occipital ERP traces. Thus, the same topography at different time points can be observed only if higher areas provide feedback to early visual cortex for representing second-order objects.

DISCUSSION

Location is a fundamental feature of object representation. We used multivariate pattern classification techniques to trace the temporal evolution of coarse location representation of various objects. We found that location information is present early for first-order objects and progressively later for more complex objects. Despite these differences in the timing, location was represented similarly in the brain across various objects. That is, the same coarse location information is available to the brain, particularly in early visual areas, for all tested object types, albeit at different time points. Finally, the estimated sources of the relevant scalp activity of all object types were situated in early visual cortex and showed considerable overlap. These data strongly suggest a role for feedback in representing locations. The finding that human and classifier performance are tightly coupled although the tasks are orthogonal to each other suggests that both utilize the same representation for their respective tasks. It can be argued that the visual system, parsimoniously, creates a

single bound representation that includes both identity and location information (contrary to the two-stream hypothesis of visual processing), and this single representation can be drawn upon for various tasks, as required.

These findings are in line with studies that suggest a role for early visual cortex in segregating second order textures (Super, Spekreijse, & Lamme, 2003; Super et al., 2001b; Lamme et al., 1999; Zipser et al., 1996). These studies have posited two stages of processing textures: (1) an early component in the striate cortex that allows detection of the texture borders in stimuli such as the one used in the current study, and (2) a second, later stage of contextual modulation of activity in the striate cortex that is associated with segregation of the texture surface and filling in of the features. This stage is thought to require feedback from higher visual areas that are presumably involved in texture processing, such as V4 and LOC (Thielscher et al., 2008; Larsson et al., 2006; Kastner et al., 2000). Until recently, it was unclear whether this modulation of V1 activity was causally involved in texture segregation or whether it was an epiphenomenon of texture processing in higher visual areas. Recent studies using TMS have provided evidence that lack of feedback to V1 at the second stage of processing reduces the ability to segregate textures (Wokke et al., 2012). However, it is possible that coarse location encoding of textures could occur in higher visual areas without the need for feedback for two reasons: (1) border detection occurs in the feed-forward sweep and might be sufficient to assign coarse position to the target (Leventhal, Wang, Schmolesky, & Zhou, 1998) and indeed might be the antecedent to targeted feedback to specific retinotopic locations in V1 and (2) although contextual modulation of V1 activity through feedback might enhance texture segregation, as shown by TMS and cortical cooling studies (Wokke et al., 2012; Hupe et al., 1998, 2001), it might not be necessary or relevant for coarse location processing. Nevertheless, our study shows that even coarse location coding depends on feedback processing to early visual areas.

Although we were mainly interested in examining the neural processes underlying coarse location coding, we also tested if we could determine whether EEG signals contained information about the objects' shape. However, the classifiers were unsuccessful at predicting object shapes. This null result could be because of several factors, including peripheral presentation of the stimuli, insensitivity of EEG signals to shape coding, among others. Because it was not our primary question, we did not pursue this thread of inquiry.

Alternative Explanations

1. *Adjacent areas hypothesis*: Although the feedback hypothesis best explains our results—particularly, the finding that coarse location information has the same topography for all objects, but at different delays—it is possible that a feed-forward explanation might still

account for the data. Consider the scenario: First-order stimuli are processed by early visual cortex, say the striate cortex, and each of the second-order stimuli are processed in successive, adjacent, downstream regions in the extrastriate cortex, say V2 to V4. These areas are adjacent enough that the poor spatial resolution of EEG would not be able to differentiate between them, and the different timings can be explained in terms of corresponding delays in processing of the second-order textures. Thus, the topography would be the same or very similar but occurring at different time points, as observed here. However, this explanation is unlikely for three reasons.

- a. According to this version of the feed-forward hypothesis, the peak classifier of each object (as seen in Figure 2A) reflects the latency and topography of the area in which that object is segregated (say, V1, V2, and V4 for successively more complex objects). Because the areas are close to each other, the topographies for the various objects should be very similar, but not identical to each other. Consequently, if the peak classifiers of complex, second-order objects were applied to predicting the coarse location of first-order objects, they should do an adequate job. However, when decoding the location of first-order objects, the best topographical match for these second-order object classifiers should take place when the first-order object signals reach higher areas (namely, the areas involved in segregating second-order objects, say, V2 and V4). Therefore, the peak latencies of predictions should also be correspondingly later (they should peak at the peak latencies of the second-order object classifiers). As predicted, the peak classifiers of second-order objects can classify the coarse location of high-luminance objects (first-order), and this decoding is not as accurate as that by first-order classifiers (Figure 3B). But contrary to the feed-forward hypothesis, the peak latencies are not shifted to later time points for the second-order classifiers but occur at the same latencies as those for first-order classifiers. The same latency for all peak classifiers when decoding the location of high-luminance objects suggests that the topography on which these classifiers are based on are not only similar but identical, implying feedback processes.
- b. A feed-forward sweep, without any feedback processing, is too fast to account for the latencies observed. Humans can segregate target objects and categorize them as quickly as 230 msec (Thorpe & Fabre-Thorpe, 2001; Thorpe et al., 1996). Similarly, saccadic latencies in humans are even faster—as short as 120 msec (Kirchner & Thorpe, 2006). Furthermore, multiple studies have documented very short neuronal latencies at various stages of the visual pathway. The median latencies of neurons in the visual cortex start at around 60 msec for V1 and reach

the “top” of the hierarchy, the inferotemporal cortex, within 120–150 msec after stimulus onset, with each intermediate stage taking an additional 10–20 msec longer than its preceding stage (Hegde & Van Essen, 2004, 2006; Lamme & Roelfsema, 2000; Schmolesky et al., 1998; Bullier, Hupe, James, & Girard, 1996; Nowak et al., 1995; Maunsell & Gibson, 1992; Raiguel, Lagae, Gulyas, & Orban, 1989). Neuronal latencies are slightly longer but comparable in humans, with latencies in the temporal cortex around 150 msec (Yoshor, Bosking, Ghose, & Maunsell, 2007). These neural and behavioral findings suggest that the feed-forward sweep, even with its complex computation, including segmentation, is probably completed within the first 200 msec. Given this rapid processing in the feed-forward sweep, it is unlikely that the large latency differences observed here between first- and second-order stimuli (latency differences were as much as 150 msec and peak latencies for texture objects were around 250+ msec) can be purely feed-forward in nature, particularly if they are constrained to be processed in adjacent areas, as proposed by the above hypothesis. We feel that the data strongly favors the feedback hypothesis.

- c. Finally, although source localization based on EEG activity is not precise, the finding that the sources of all object types overlapped considerably and were situated in early visual areas supports the argument that the adjacent area hypothesis cannot be sustained.
2. *Weak stimulus hypothesis:* A second related alternative hypothesis is that the location of all objects are processed in the same visual area (and hence provide the same topography for the classifier), but because the second-order stimuli are, in some sense, “weaker” stimuli for that area, they are processed later (longer latencies) and not as well (lower accuracies), just as a low-contrast stimulus is processed later than a high-contrast stimulus in V1 (Gawne et al., 1996; Lennie, 1981). Once again, this hypothesis does not explain the current data because (1) it is unclear how the second-order stimuli can “weakly” drive any part of the visual cortex, because they are made up of high-contrast (and high-luminance) components; (2) edge detection seems to happen equally rapidly in the early visual cortex for texture defined stimuli as for contrast defined stimuli (Leventhal et al., 1998; Chaudhuri & Albright, 1997), which could potentially be used for location representation. It has been shown that an object can be detected rapidly even when its shape cannot be discriminated (Bowers & Jones, 2008; Mack, Gauthier, Sadr, & Palmeri, 2008), suggesting that location information of complex stimuli is available even if other aspects of the stimulus are not processed; (3) there were no differences in latencies or accuracy between high- and low-luminance first-order stimuli,

although there were clear luminance differences, and hence possibly latency differences, between the two stimulus types; and (4) even if the second-order stimuli were weak drivers of certain visual areas where they are processed, the latency difference between first- and second-order stimuli are too large for the two stimuli to be processed in the same region, without some sort of feedback sustaining the neural activity. Given the high-contrast nature of the components of all second-order stimuli, the signals elicited by them would presumably reach any given visual area around the same time as first-order stimuli. Thus latency differences in classifier performance would have to be because of “processing time” differences at the same neural site. This cannot occur without some sort of feedback activity (local or top-down) sustaining the processing of the second-order stimuli for a longer duration than for first-order stimuli at the relevant site. Thus, we argue that these processing time delays are because of differences in the need for feedback processing and unlikely to be because of the “weakness” of the stimuli.

Locus of Location Information

We have claimed that coarse location coding occurs in the early visual cortex, based on the early onset latencies of classification performance and the higher weighting given to the parieto-occipital electrodes by the classifier. However, because of the poor spatial resolution of EEG, one cannot be certain of the precise site of location coding, even considering our source localization results. However, these findings are further supported by recent findings in texture processing based on intracranial recordings and TMS (Wokke et al., 2012; Super & Lamme, 2007; Super et al., 2001b).

Recent studies have suggested that coarse location could be encoded in high-level visual areas such as the LOC, FFA, and PPA. Hence, it is possible that the visual system utilizes this information when coarse localization is sufficient for a task. Our findings that coarse location representation is the same across all object types suggest otherwise. We found that even in circumstances where coarse location information would suffice, location is represented most reliably, through feedback, in the coordinates afforded by the early retinotopic cortex.

However, this raises the related question of the neural representation of precise locations of complex objects. Because higher visual areas have not been shown to be retinotopic or to have precise location information (but see Fischer, Spotswood, & Whitney, 2011; Larsson & Heeger, 2006), precise localization must rely on communication with areas that possess such information. These areas might be the early visual cortex, as for coarse localization, or regions of the “where” pathway that process spatial information (e.g., see Bremmer, Schlack, Duhamel, Graf, & Fink, 2001; Andersen, Essick, & Siegel, 1985). If the dorsal pathway can compute object location without

any input from the ventral stream, one would expect that location information would be identical across various object types, as we found, but also that this representation would be evoked at the same time point, unlike what we found. Hence, it does not seem plausible that neither coarse nor precise location is represented in purely dorsal areas. However, further studies are necessary to test this inference. On the other hand, although we might be tempted to conclude that a feedback communication loop between higher (ventral) and early visual areas supports location representation regardless of task requirements, our study cannot support this assertion as we did not test precise localization. Nevertheless, we feel that the methodology developed here is capable of disentangling these possibilities.

Applications of Classification Methods to the Temporal Domain

We have shown that classification methods can be an effective way to probe the evolution of neural representations of visual objects. In the past, classification methods have been applied predominantly to fMRI and neurophysiological data in cognitive neuroscience and also in the brain-computer interface literature, where the goal is to classify scalp electrical patterns to signal a choice, without a behavioral response from the observer (e.g., Friedrich, Scherer, & Neuper, 2012; Green & Kalaska, 2011; Mak et al., 2011; Palaniappan, Syan, & Paramesran, 2009; Mensh, Werfel, & Seung, 2004). This technique has been rarely applied to electrophysiological measures such as EEG and MEG to study the temporal evolution of cognitive processes underlying perception. It is to be noted that some studies have used this method to assess cognitive processes, but they were not interested in extracting the detailed temporal dynamics of the processes (Chan, Halgren, Marinkovic, & Cash, 2011; Murphy et al., 2011). However, a few recent studies have begun to address this lack (Carlson, Hogendoorn, & Kanai, et al., 2011, 2013). For example, a recent study (Carlson, Hogendoorn, & Kanai, et al., 2011) showed that MEG responses could be used to categorize objects such as faces and cars. Furthermore, the location of such objects could also be decoded, and the precision of this decoding depended on their cortical separation, arguing for representations in the early visual cortex, consistent with current results. We suggest that such decoding methods are useful in determining the temporal evolution of information transfer during the course of cognitive processing, given the high temporal resolution and the breadth of spatial coverage of techniques like EEG and MEG.

Another advantage of using the classification technique to assess cognitive or visual processes is that, unlike electrophysiological recordings or even TMS studies, it is completely noninvasive. For example, we can obtain strong evidence for feedback without needing to disrupt any ongoing processes and thus assess processes in their “natural” state to the extent possible.

Conclusion

The current study explored the nature and temporal evolution of coarse location representation using multivariate pattern classification techniques, along with complementary source localization analysis. We found that coarse location of an object, irrespective of its complexity, is represented in early visual areas. The nature of this representation is the same for all tested object types, with differences in their signal-to-noise ratios, but manifests at different time points, depending on where the object is processed in the ventral visual stream. These findings support the feedback hypothesis of coarse location representation.

Acknowledgments

This project was supported by a EURYI award grant to R. V. We thank an anonymous reviewer for suggesting the source localization analysis, which provided an independent test of the hypotheses.

Reprint requests should be sent to Ramakrishna Chakravarthi, Department of Psychology, University of Aberdeen, William Guild Building, Kings College, Old Aberdeen, United Kingdom, AB24 3FX, or via e-mail: rama@abdn.ac.uk.

REFERENCES

- Aggelopoulos, N. C., & Rolls, E. T. (2005). Scene perception: Inferior temporal cortex neurons encode the positions of different objects in the scene. *European Journal of Neuroscience*, *22*, 2903–2916.
- Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, *230*, 456–458.
- Bowers, J. S., & Jones, K. W. (2008). Detecting objects is easier than categorizing them. *Quarterly Journal of Experimental Psychology*, *61*, 552–557.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.
- Bremmer, F., Schlack, A., Duhamel, J. R., Graf, W., & Fink, G. R. (2001). Space coding in primate posterior parietal cortex. *Neuroimage*, *14*, S46–S51.
- Bullier, J. (2001). Integrated model of visual processing. *Brain Research. Brain Research Reviews*, *36*, 96–107.
- Bullier, J., Hupe, J. M., James, A., & Girard, P. (1996). Functional interactions between areas V1 and V2 in the monkey. *Journal of Physiology*, *90*, 217–220.
- Carlson, T., Hogendoorn, H., Fonteijn, H., & Verstraten, F. A. (2011). Spatial coding and invariance in object-selective cortex. *Cortex*, *47*, 14–22.
- Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 msec. *Journal of Vision*, *13*, 1–19.
- Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution decoding of object position and category. *Journal of Vision*, *11*.
- Cavanagh, P., Hunt, A. R., Afraz, A., & Rolfs, M. (2010). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences*, *14*, 147–153.
- Chan, A. M., Halgren, E., Marinkovic, K., & Cash, S. S. (2011). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *Neuroimage*, *54*, 3028–3039.
- Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*, 1–27.
- Chaudhuri, A., & Albright, T. D. (1997). Neuronal responses to edges defined by luminance vs. temporal texture in macaque area V1. *Visual Neuroscience*, *14*, 949–962.
- Cichy, R. M., Chen, Y., & Haynes, J. D. (2011). Encoding the identity and location of objects in human LOC. *Neuroimage*, *54*, 2297–2307.
- Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*, 16.1–17.
- d’Avossa, G., Tosetti, M., Crespi, S., Biagi, L., Burr, D. C., & Morrone, M. C. (2007). Spatiotopic selectivity of BOLD responses to visual motion in human area MT. *Nature Neuroscience*, *10*, 249–255.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.
- 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.
- DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*, 415–434.
- Drewes, J., Trommershauser, J., & Gegenfurtner, K. R. (2011). Parallel visual search and rapid animal detection in natural scenes. *Journal of Vision*, *11*, 1–21.
- Edelman, S., & Intrator, N. (2000). (Coarse coding of shape fragments) + (retinotopy) approximately = representation of structure. *Spatial Vision*, *13*, 255–264.
- Essen, D. C., & Zeki, S. M. (1978). The topographic organization of rhesus monkey prestriate cortex. *Journal of Physiology*, *277*, 193–226.
- Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*, 1–29.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1–47.
- Fischer, J., Spotswood, N., & Whitney, D. (2011). The emergence of perceived position in the visual system. *Journal of Cognitive Neuroscience*, *23*, 119–136.
- Friedrich, E. V., Scherer, R., & Neuper, C. (2012). The effect of distinct mental strategies on classification performance for brain-computer interfaces. *International Journal of Psychophysiology*, *84*, 86–94.
- Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *Journal of Neuroscience*, *28*, 3988–3999.
- Gawne, T. J., Kjaer, T. W., & Richmond, B. J. (1996). Latency: Another potential code for feature binding in striate cortex. *Journal of Neurophysiology*, *76*, 1356–1360.
- Golomb, J. D., & Kanwisher, N. (2012). Higher level visual cortex represents retinotopic, not spatiotopic, object location. *Cerebral Cortex*, *22*, 2794–2810.
- Green, A. M., & Kalaska, J. F. (2011). Learning to move machines with the mind. *Trends in Neurosciences*, *34*, 61–75.
- Hegde, J., & Van Essen, D. C. (2004). Temporal dynamics of shape analysis in macaque visual area v2. *Journal of Neurophysiology*, *92*, 3030–3042.
- Hegde, J., & Van Essen, D. C. (2006). Temporal dynamics of 2D and 3D shape representation in macaque visual area V4. *Visual Neuroscience*, *23*, 749–763.
- Heinen, K., Jolij, J., & Lamme, V. A. (2005). Figure-ground segregation requires two distinct periods of activity in

- V1: A transcranial magnetic stimulation study. *NeuroReport*, *16*, 1483–1487.
- 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.
- Hesse, C., Ball, K., & Schenk, T. (2012). Visuomotor performance based on peripheral vision is impaired in the visual form agnostic patient DF. *Neuropsychologia*, *50*, 90–97.
- 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.
- Hupe, J. M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., & Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. *Journal of Neurophysiology*, *85*, 134–145.
- Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, *394*, 784–787.
- Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology*, *73*, 218–226.
- Ives, H. E., & Kingsbury, E. F. (1914). LXXIV. The theory of the flicker photometer. *Philosophical Magazine Series 6*, *28*, 708–728.
- Juan, C. H., & Walsh, V. (2003). Feedback to V1: A reverse hierarchy in vision. *Experimental Brain Research*, *150*, 259–263.
- Kastner, S., De Weerd, P., & Ungerleider, L. G. (2000). Texture segregation in the human visual cortex: A functional MRI study. *Journal of Neurophysiology*, *83*, 2453–2457.
- Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*, 1762–1776.
- Kleiner, M., Brainard, D., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(*ECVP Abstract Suppl.*), 14.
- Lamme, V. A., Rodriguez-Rodriguez, V., & Spekreijse, H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cerebral Cortex*, *9*, 406–413.
- Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*, 571–579.
- Lamme, V. A., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *Journal of Cognitive Neuroscience*, *14*, 1044–1053.
- Larsson, J., & Heeger, D. J. (2006). Two retinotopic visual areas in human lateral occipital cortex. *Journal of Neuroscience*, *26*, 13128–13142.
- Larsson, J., Landy, M. S., & Heeger, D. J. (2006). Orientation-selective adaptation to first- and second-order patterns in human visual cortex. *Journal of Neurophysiology*, *95*, 862–881.
- Lennie, P. (1981). The physiological basis of variations in visual latency. *Vision Research*, *21*, 815–824.
- Leventhal, A. G., Wang, Y., Schmolesky, M. T., & Zhou, Y. (1998). Neural correlates of boundary perception. *Visual Neuroscience*, *15*, 1107–1118.
- Li, P., Zhu, S., Chen, M., Han, C., Xu, H., Hu, J., et al. (2013). A motion direction preference map in monkey V4. *Neuron*, *78*, 376–388.
- 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.
- Livingstone, M. S., & Hubel, D. S. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, *240*, 740–749.
- Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex*, *5*, 270–288.
- Macevoy, S. P., & Epstein, R. A. (2007). Position selectivity in scene- and object-responsive occipitotemporal regions. *Journal of Neurophysiology*, *98*, 2089–2098.
- Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review*, *15*, 28–35.
- Mak, J. N., Arbel, Y., Minett, J. W., McCane, L. M., Yuksel, B., Ryan, D., et al. (2011). Optimizing the P300-based brain-computer interface: Current status, limitations and future directions. *Journal of Neural Engineering*, *8*, 025003.
- Maunsell, J. H. (1987). Physiological evidence for two visual subsystems. In L. M. Vaina (Ed.), *Matters of intelligence* (pp. 59–87). Dordrecht: Reidel Publishing Company.
- Maunsell, J. H., Ghose, G. M., Assad, J. A., McAdams, C. J., Boudreau, C. E., & Noerager, B. D. (1999). Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. *Visual Neuroscience*, *16*, 1–14.
- Maunsell, J. H., & Gibson, J. R. (1992). Visual response latencies in striate cortex of the macaque monkey. *Journal of Neurophysiology*, *68*, 1332–1344.
- Melcher, D., & Morrone, M. C. (2003). Spatiotopic temporal integration of visual motion across saccadic eye movements. *Nature Neuroscience*, *6*, 877–881.
- Mensh, B. D., Werfel, J., & Seung, H. S. (2004). BCI competition 2003-data set Ia: Combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals. *IEEE Transactions on Biomedical Engineering*, *51*, 1052–1056.
- Milner, A. D., & Goodale, M. A. (1993). Visual pathways to perception and action. *Progress in Brain Research*, *95*, 317–337.
- Murphy, B., Poesio, M., Bovolenta, F., Bruzzone, L., Dalponte, M., & Lakany, H. (2011). EEG decoding of semantic category reveals distributed representations for single concepts. *Brain and Language*, *117*, 12–22.
- Nowak, L. G., Munk, M. H., Girard, P., & Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Visual Neuroscience*, *12*, 371–384.
- Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155PB*, 23–36.
- Op De Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology*, *426*, 505–518.
- Palaniappan, R., Syan, C. S., & Paramesran, R. (2009). *Current practices in electroencephalogram-based brain-computer interfaces* (2nd ed.). Hershey, PA: IGI Global.
- Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, *292*, 510–512.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Raiguel, S. E., Lagae, L., Gulyas, B., & Orban, G. A. (1989). Response latencies of visual cells in macaque areas V1, V2 and V5. *Brain Research*, *493*, 155–159.
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87.

- Ro, T., Breitmeyer, B., Burton, P., Singhal, N. S., & Lane, D. (2003). Feedback contributions to visual awareness in human occipital cortex. *Current Biology*, *13*, 1038–1041.
- Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nature Neuroscience*, *7*, 982–991.
- 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.
- Schenk, T., & McIntosh, R. D. (2009). Do we have independent visual streams for perception and action? *Cognitive Neuroscience*, *1*, 52–62.
- Schiller, P. H., & Logothetis, N. K. (1990). The color-opponent and broad-band channels of the primate visual system. *Trends in Neurosciences*, *13*, 392–398.
- Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., et al. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology*, *79*, 3272–3278.
- Schwartz, E. L., Desimone, R., Albright, T. D., & Gross, C. G. (1983). Shape recognition and inferior temporal neurons. *Proceedings of the National Academy of Sciences, U.S.A.*, *80*, 5776–5778.
- Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., & Poggio, T. (2007). A quantitative theory of immediate visual recognition. *Progress in Brain Research*, *165C*, 33–56.
- Serre, T., & Poggio, T. (2010). A neuromorphic approach to computer vision. *Communications of the ACM*, *53*, 54–61.
- Super, H., & Lamme, V. A. (2007). Altered figure-ground perception in monkeys with an extra-striate lesion. *Neuropsychologia*, *45*, 3329–3334.
- Super, H., Spekreijse, H., & Lamme, V. A. (2001a). A neural correlate of working memory in the monkey primary visual cortex. *Science*, *293*, 120–124.
- Super, H., Spekreijse, H., & Lamme, V. A. (2001b). Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nature Neuroscience*, *4*, 304–310.
- Super, H., Spekreijse, H., & Lamme, V. A. (2003). Figure-ground activity in primary visual cortex (V1) of the monkey matches the speed of behavioral response. *Neuroscience Letters*, *344*, 75–78.
- Thielscher, A., Kollé, M., Neumann, H., Spitzer, M., & Gron, G. (2008). Texture segmentation in human perception: A combined modeling and fMRI study. *Neuroscience*, *151*, 730–736.
- Thorpe, S. J., & Fabre-Thorpe, M. (2001). Neuroscience. Seeking categories in the brain. *Science*, *291*, 260–263.
- Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences*, *11*, 58–64.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (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.
- Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology*, *51*, 167–194.
- Wokke, M. E., Sligte, I. G., Steven Scholte, H., & Lamme, V. A. (2012). Two critical periods in early visual cortex during figure-ground segregation. *Brain Behavior*, *2*, 763–777.
- Yoshor, D., Bosking, W. H., Ghose, G. M., & Maunsell, J. H. (2007). Receptive fields in human visual cortex mapped with surface electrodes. *Cerebral Cortex*, *17*, 2293–2302.
- Zipser, K., Lamme, V. A., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *Journal of Neuroscience*, *16*, 7376–7389.
- Zoccolan, D., Kouh, M., Poggio, T., & DiCarlo, J. J. (2007). Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *Journal of Neuroscience*, *27*, 12292–12307.