Representational dynamics of object vision: The first 1000 ms

Thomas Carlson — Department of Cognitive Sciences, Macquarie University, Sydney, NSW, Australia; Centre for Cognition & its Disorders, Macquarie University, Sydney, NSW, Australia; Department of Psychology, University of Maryland, College Park, MD, USA

David A. Tovar — Department of Psychology, University of Maryland, College Park, MD, USA

Arjen Alink — MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom

Nikolaus Kriegeskorte — MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom

Human object recognition is remarkably efficient. In recent years, significant advancements have been made in our understanding of how the brain represents visual objects and organizes them into categories. Recent studies using pattern analyses methods have characterized a representational space of objects in human and primate inferior temporal cortex in which object exemplars are discriminable and cluster according to category (e.g., faces and bodies). In the present study we examined how category structure in object representations emerges in the first 1000 ms of visual processing. In the study, participants viewed 24 object exemplars with a planned categorical structure comprised of four levels ranging from highly specific (individual exemplars) to highly abstract (animate vs. inanimate), while their brain activity was recorded with magnetoencephalography (MEG). We used a sliding time window decoding approach to decode the exemplar and the exemplar's category that participants were viewing on a moment-to-moment basis. We found exemplar and category membership could be decoded from the neuromagnetic recordings shortly after stimulus onset (<100 ms) with peak decodability following thereafter. Latencies for peak decodability varied systematically with the level of category abstraction with more abstract categories emerging later, indicating that the brain hierarchically constructs category representations. In addition, we examined the stationarity of patterns of activity in the brain that encode object category information and show these patterns vary over time, suggesting the brain might use flexible time varying codes to represent visual object categories.

Introduction

Human object recognition is remarkably efficient, taking less than a fraction of second. In recent years, great strides have been made in our understanding of the neural machinery that underlies this remarkable capacity in humans. Functional magnetic resonance imaging (fMRI) enabled the study of coarse scale organization of object representations in the ventral temporal pathway, the region of the brain that underlies object vision (Logothetis & Sheinberg, 1996; Ungerleider & Miskin, 1982). Early studies identified regions in the ventral occipital temporal (VOT) cortex that respond selectively to objects (Malach et al., 1995), and a few specific categories, in particular faces, places, and bodies (Downing, Jiang, Shuman, & Kanwisher, 2001; Epstein & Kanwisher, 1998; Kanwisher, McDermott, & Chun, 1997). Studies using multivariate pattern analysis (MVPA), which arguably give researchers access to more fine scale organization (Kamitani & Tong 2005), challenged the notion that there are regions in VOT that are selective for categories (Haxby et al., 2001), instead arguing for a large scale distributed coding scheme.

Citation: Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision, 13*(10):1, 1–19, http://www.journalofvision.org/content/13/10/1, doi:10.1167/13.10.1.

doi: 10.1167/13.10.1 Received January 16, 2013; published August 1, 2013 ISSN 1534-7362

Recent work utilizing MVPA has revealed a *representational space* of objects in monkey and human inferior temporal cortex (IT) (Kiani, Esteky, Mirpour, & Tanaka, 2007; Kriegeskorte et al., 2008), in part reconciling these disparate views. In this space, object representations form a hierarchy of clusters that reflect conventional object categories, while discriminating exemplars within categories.

How does this representational space emerge? The aforementioned fMRI studies focused on the information in spatial brain activity patterns, derived from neuronal activity averaged across time. Noninvasive electrophysiological studies in humans have investigated the temporal structure of visual object responses, averaged across space (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Liu, Harris, & Kanwisher, 2002). While this approach is successful at revealing certain category-related overall responses, it cannot address the emergence of the detailed hierarchical clusters in IT representational space. Monkey studies investigating the temporal structure of visual object responses have similarly neglected the information in spatial patterns across populations of neurons, reporting, for example, that the response latency of neurons can vary between object categories (Bell et al., 2011; Kiani, Esteky, & Tanaka, 2005). One notable study using a sliding time window decoding approach to investigate the decodability of categories and exemplars from monkey IT population response patterns found that category and exemplar information appeared at similar latencies: about 100 ms after stimulus onset (Hung, Kreiman, Poggio, & DiCarlo, 2005). Other neurophysiological studies, however, have reported discrepancies in the emergence of category and exemplar representations dependent on the level categorical specificity (Matsumoto, Okada, Sugase-Miyamoto, Yamane, & Kawano, 2005; Sugase, Yamane, Ueno, & Kawano, 1999). It thus remains unclear if, and if so how, categorical structure emerges in the brain.

In recent years, neurophysiological investigations have moved towards studying the dynamics of population codes over time (Buonomano & Maass, 2009; Crowe, Averbeck, & Chafee, 2010; Mazor & Laurent 2005; Nikolic, Hausler, Singer, & Maass, 2009; Rabinovich, Huerta, & Laurent, 2008) to study how information is encoded in these representations. In the brief time it takes for a human to recognize an object, the brain will rapidly undergo changes in its representational state to promote recognition. Visual areas represent objects as features (e.g., oriented edges and color) shortly after the onset of a stimulus. These representations later will be refined into object representations in higher visual areas. MEG and EEG measure whole brain activity with millisecond temporal resolution—the brain's representational state at a given time—which affords researchers the capacity to study the representational dynamics of perceptual processes. Contemporary human neuroimaging, however, has largely neglected to examine how these states emerge and to characterize their dynamics (see Philiastides & Sajda, 2006).

We have previously employed a sliding time window MEG decoding approach to show position invariant (or position tolerant) object category information emerges shortly after the onset of the visual stimulus (Carlson, Hogendoorn, Kanai, Mesik, & Turret, 2011). In the present study, we extend this work to study the emergence of object representations in the human brain. In particular, we asked (1) when do exemplar representations of several categories become discriminable, (2) when do exemplar representations cluster in contiguous regions corresponding to the categories, (3) when do decodable category divisions emerge, and (4) what are the temporal characteristics of the patterns of brain activity that enable decoding? To address these questions, we used magnetoencephalography (MEG) to record brain activity in high temporal resolution while participants' viewed 24 object exemplars with a planned categorical structure; and, to characterize the structure of object representations on a moment-to-moment basis we used sliding-temporal-window pattern decoding.

Methods

Participants

Twenty volunteers (seven male, 13 female) with an average age of 19.9 years participated in the experiment. All participants had normal or corrected-to-normal vision. Informed written consent was obtained from each volunteer prior to the experiment. Participants were paid a base rate of $15 for their participation. They were awarded an additional $5 for achieving a high level of performance on the experimental task (see below). The University of Maryland Institutional Review Board approved all experimental procedures.

Display apparatus

Subjects were supine in the recording chamber. Stimuli were projected onto a translucent screen located 30 cm above the participant. Experiments were run on a Dell PC desktop computer using MATLAB (Natick, MA).

Stimuli

Stimuli were a set of 24 naturalistic images of objects segmented from personal photos, and stock photos

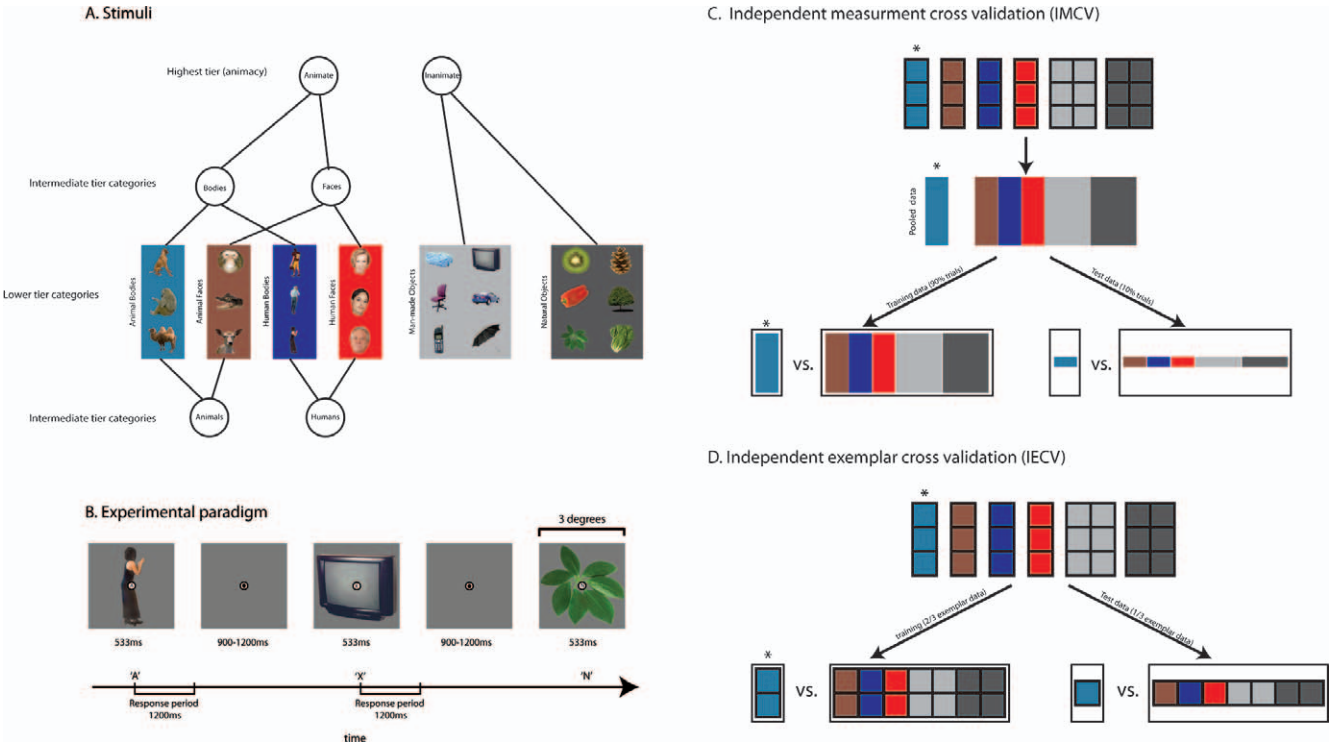


Figure 1. Experimental methods. (A) Stimuli. Stimuli were 24 images of objects with a planned hierarchical category structure. The exemplar level is the individual images. These images group into six lower tier categories: animal bodies, animal faces, human bodies, human faces, man-made objects, and natural objects. The animate objects group into images depicting intermediate tier categories faces and bodies. They also can be grouped as images depicting intermediate tier categories humans and animals (Level 2b). The highest tier distinguishes animate and inanimate objects. (B) Trial sequence. Each image was displayed for 533 ms with a variable interstimulus interval that ranged from 900 to 1200 ms. In the center of each image was a letter. The participants' task was to report whether the letter was a consonant or vowel. (C) Same exemplar decoding. The classifier is trained to decode the category of the stimulus. 90% of the data is used to train the classifier. The trained classifier is then tested with the remaining 10% of the data. Color coding in the figure corresponds to the colors in Figure 1A. In the example shown, the classifier is trained to decode whether or not the image shown to the observer is an animal bodies (denoted by the asterisk). (D) Novel exemplar decoding. The classifier is again trained to classify animal bodies (denoted by the asterisk) from other stimuli. Here, the classifier is trained with block of exemplar data. The ratio of training to test is 2:1. Two thirds of the exemplars from each category are used to train the classifier. The excluded exemplars' data are used to test the classifier.

from www.photos.com and www.gettyimages.com. The objects in the images had a planned categorical structure informed by earlier studies (Kiani et al., 2007; Kriegeskorte et al., 2008), which observed clustering of object exemplars by object category in IT cortex, in particular faces and bodies, and a broad category distinction between animate and inanimate objects. The set of object exemplars used in the present study were selected such that they could be similarly hierarchically grouped (Figure 1A). The hierarchy has an exemplar level, which are the individual images and several higher order groupings (i.e., categories). The highest tier in hierarchy (animacy) was balanced such that there were an equal number of animate (12 exemplars) and inanimate objects (12 exemplars). Below this, the animate objects grouped into four intermediate tier categories. The first two intermediate tier categories were faces (six exemplars) and bodies (six exemplars), categories that have delineated regions of the human brain that show selectivity for the category (Downing et al., 2001; Kanwisher et al., 1997). The other two intermediate tier categories are human and animal, which are made of the same set of images, but grouped differently. The lowest tier categories grouped the exemplars into animal bodies (three exemplars), animal faces (three exemplars), human bodies (three exemplars), human faces (three exemplars), man-made objects (six exemplars), and natural occurring objects (six exemplars).

Visual models of the stimuli

The use of naturalistic images introduces the possibility that low-level feature differences can dis-

criminate categories (e.g., based on shape and color) and account for the decoding results. We employed several visual feature models to examine this possibility. The first model compares the difference between the shapes of the figures (i.e., the silhouettes; Jaccard, 1901). The second compares the difference between the images in uniform color space (CIE) color space. The third compares the images using a hierarchal model of visual object processing (HMAX; Riesenhuber & Poggio, 1999). In particular, we compared the image representations in the C2 layer of HMAX. For each model, we constructed a dissimilarity matrix (DSM) representing the difference between the individual images according the model outputs. Each model's DSM was constructed by correlating the model outputs for all possible pairwise comparisons between the images.

Experimental design

Figure 1B diagrammatically shows a sequence of trials. In each trial, an image of an object was shown for 533 ms. The inter-trial interval between the stimuli was random (uniform distribution) with a range of 900 to 1200 ms. For the purpose of the task, a small letter was superimposed onto the center image. The letter was randomly selected from the following set: (O, U, R, N, X, S, T).

Participants were shown blocks of trials, composed of a sequence of 240 images. Each exemplar was shown ten times in each block (24 Exemplars $\times$ 10 = 240 Trials). The order of the images was randomized for each block. Participants were requested to perform eight blocks of trials. Seventeen participants completed the entire experiment (eight blocks, 80 trials per exemplar); one terminated the experiment after seven blocks (seven blocks, 70 trials per exemplar); one terminated the experiment after six blocks (six blocks, 60 trials per exemplar); and one subject opted to complete 10 blocks (10 blocks, 100 trials per exemplar).

Experimental task/results

Participants performed a task unrelated to the aims of the experiment to encourage them to maintain vigilance. The task was to report as quickly and accurately as possible whether the letter at fixation was a vowel or consonant. Participants received feedback after each trial. After each block, they received a summary of their performance for the block. Additional monetary compensation was awarded to participants with an average reaction time less than 500 ms and accuracy above 95% correct for the entire experiment. The mean accuracy across participants was 93% correct (standard deviation 5.1%). The average reaction time was 453 ms (standard deviation 48 ms).

MEG recordings and data preprocessing

MEG recordings were made with a 160 channel whole-head axial gradiometer (KIT, Kanazawa, Japan). Signals were digitized at 1000 Hz and filtered online from 0.1 to 200 Hz using first-order RC filters. Offline, time-shifted principal component analysis (TSPCA) was used to denoise the data (de Cheveigne & Simon, 2007). TSPCA removes noise from neurophysiological signals by making use of hardware reference channels that measure environmental noise. TSPCA filters the reference channels to optimally estimate the noise in the signal channels, using PCA to generate the filters, and then subtracts the estimated noise from the signals.

Trials were epoched from 0.1 s before to 1 s after stimulus onset. Each trial was visually inspected for eye-movement artifacts. The average rejection rate was 7.4% with a standard deviation of 2.7% across participants. Principle component analysis (PCA) was used to reduce the dimensionality of the dataset. We used a criterion of retaining 98% of the variance, which on average reduced the dimensionality of the dataset from 157 recording channels to 58 components (standard deviation 4.6 components). The time series data was resampled to 50 Hz. The choice of sampling rate was selected to balance several considerations. Ideally, we would like to retain as much temporal resolution as possible, which favors the use of the original sampling rate (1000 Hz). At the same time, a lower sampling rate reduces the time for processing the data and increases the signal to noise. Our preliminary analyses using sampling rates of 1000, 200, 50, and 20 Hz found 50 Hz to be a good balance of these factors. The data was downsampled to 50 Hz in MATLAB using function that low-pass filters the data before downsampling. The latency latency offset introduced by this filter (estimated by simulation to be $\sim$20 ms; see VanRullen, 2011) was corrected for after downsampling.

Pattern classification analysis

The aim of the study was to examine when and how categorical structure emerges in the brain. To this end, we used naïve Bayes implementation of linear discriminant analysis (LDA, Duda, Hart, & Stork, 2001) to do single trial classification of the exemplar and category of the stimuli that participants were viewing. Additional analyses were conducted using Euclidean distance and correlation based decoding methods.

LDA accuracy was found to be consistently higher. Only the LDA results are presented. Classification results are reported in terms of d prime.

Sliding-time-window analysis

A new classifier was trained and tested for each time point. This sliding time window approach was used to study the emerging categorical structure of object representations using cross validation approaches:

Independent measurement cross validation (IMCV, Figure 1C)

In this analysis, we trained a classifier to discriminate response patterns elicited by exemplars and exemplars grouped by category. Generalization of the classifier was evaluated using on independent measurements using k-fold cross validation. We used 10-fold cross validation with a ratio of nine (training) to one (test) to evaluate decoding accuracy. In this procedure, the data is divided into two sets, which were determined by the specific comparison (e.g., human stimuli and nonhuman stimuli). The two sets of trials were subdivided into 10 subsets with individual trials assigned randomly. For each set, nine of the subsets were pooled to train the classifier (90% of the data). The remaining subsets, one from each set, were used to test the classifier (10% of the data). This procedure was repeated ten times such that each subset was tested once. Decoding results are summarized as the averaging decoding accuracies (d′) across participants.

Pairwise exemplar decoding and multidimensional scaling

For each time point, we constructed a DSM by conducting IMCV for decoding all possible pairwise exemplar comparisons. Multidimensional scaling (MDS) (Torgerson, 1958) with a metric stress criterion was used to visualize the DSM for each time point. To add continuity over time in the movie, we used a random initial position seed for MDS for the first time point. Thereafter, MDS was seeded using the solution from the preceding time point.

Category decoding using IMCV (Figure 1C)

Each category of objects was contrasted with the nonoverlapping set of exemplars from other categories using IMCV. Decoding performance was evaluated for all of the categories in each of the tiers of the stimulus hierarchy. The lower tier category contrasts were animal bodies, animal faces, human bodies, human faces, man-made objects, and natural objects. The intermediate tier category contrasts were faces, bodies, humans, and animals. The highest tier was animate versus inanimate objects.

Independent exemplar cross validation (IECV, Figure 1D)

We additionally conducted a second form of cross validation. Here, the classifier is trained and tested with different sets of *exemplars*. This analysis is important, because same-exemplar linear decoding of a category dichotomy is expected to work even for a low-level representation such as V1 that distinguishes the exemplars but does not group exemplars of each category in a contiguous region of response-pattern space. Novel-exemplar decoding serves to test for generalization of a category decoder to new exemplars. Significant novel-exemplar decodability indicates that each category is associated with a contiguous region of response-pattern space. For example in the animal face category contrast, the classifier would be trained on data from the alligator face and deer face exemplars, which would represent the target category of animal faces, and an equal proportion of exemplars from the other categories (2/3 of the exemplars) would be used to complete the training data set. The classifier would then be tested on the monkey face exemplar data, which represents the target category of animal faces, and an equal proportion of exemplars from the other categories (1/3 of the exemplars). Critically, the classifier must generalize to new exemplars to successfully decode the category of the stimulus. Some of the categories had different numbers of exemplars (e.g., three human face and six man-made objects). To maintain a balance in terms of the amount of data contributed by each category to the training and test, we used the same proportion of exemplars for training and test. Similar to same-exemplar decoding, this procedure was repeated such that each trial was tested exactly once. Note IECV preclude analyses at the exemplar level (i.e., comparing the decodability of a pair of exemplars). Pairwise decoding of exemplars therefore is limited to IMCV.

Category decoding using IECV (Figure 1D)

Category decoding was also evaluated using the IECV. The analysis was only conducted at the three categorical levels, as IECV precludes decoding at the exemplar level. The contrasts were identical to the category contrasts used in IMCV.

	Category/Visual model comparisons		
	Jaccard (rho)	HMAX (rho)	CIE (rho)
Category (lower tier)			
Animal bodies	−0.030	0.116	−0.058
Animal faces	−0.011	0.012	0.011
Human bodies	0.068	0.124	0.1761**
Human faces	0.179**	0.127	0.111
Natural objects	0.018	0.006	−0.186
Manmade objects	0.079	−0.2097**	−0.033
Category (intermediate tier)			
Faces	0.200	0.136	0.094
Bodies	−0.126	0.226	0.157
Animals	−0.024	0.153	−0.017
Humans	0.027	0.179	0.302**
Category (highest tier)			
Animacy	0.039	0.031	0.072

Table 1. Visual models and MEG category contrasts. Shown are the correlation values between three visual models of the stimuli and the category contrasts. Significance was evaluated using a bootstrap test. Significance is uncorrected for multiple comparisons. *Notes*: * indicates significance $p < 0.05$ ** indicates significance $p < 0.01$.

MEG category contrasts and low-level feature models of stimuli

To examine the relationship between the visual object categories and low-level feature differences between categories, we compared the model DSMs to models of the contrasts used in the study. For each contrast, we generated a DSM with zeros for comparisons within a category (i.e., same category) and ones for comparisons between categories (i.e., different category). For example in the human face contrast, all of the entries in the DSM corresponding to comparisons between human face stimuli would have zeros (e.g., Human Face 1 and Human Face 2), and all of the DSM entries contrasting human face with other object exemplars would be zero (e.g., Human Face 1 and the camel body). The relevant entries (i.e., those populated with zeros or ones) of the category contrast DSM were then correlated (Spearman ρ) with the corresponding entries of the visual model DSM. The resultant ρ value was compared to a null distribution of correlations values. To generate the null distribution, the visual model DSM labels were shuffled randomly and the entries of the shuffled model DSM were similarly correlated with the category contrast DSM. This procedure was repeated 1,000 times to generate a null distribution of correlation coefficients. The actual correlation coefficient was then compared to the null distribution to compute a p value.

Table 1 shows resultant correlation coefficients along with the outcome of significance tests for all the category contrasts compared to each of the models. The three models had varying degrees of success. The silhouette model could discriminate human faces from the other objects, presumably owing to the similarity in shape. HMAX could successfully discriminate the man-made objects, possibly due to man-made objects having more high contrast edges. And, the CIE model was able to capture the "human body" and "human" contrasts, possibly due to the similar coloring of human skin. Each of the models thus had some success in discriminating the categories. Even the most successful model (CIE), however, could only discriminate 20% (2 out of 10) categories analyzed in the study. In the results section, we further examine how feature differences between the categories (described by these models) might account for the decoding results.

Measures of the latency of decodability

Our study used two decodability-latency measures. *Onset latency* is the earliest time that the exemplar/category of a stimulus can be decoded from the MEG recordings. For the purposes of the present study, onset latency is defined as the earliest time point where the classifier is above chance on at least two consecutive time points (a 40 ms period) using a threshold of $p < 0.01$ (uncorrected). Significance was evaluated with a nonparametric Wilcoxon signed rank test contrasting classification performance with chance (d′ of zero). *Peak latency* is the time that the classifier maximally differentiates the exemplar/category of the stimulus. Operationally, peak latency was defined as the time with the maximum d′ value within the interval of 0 to 540 ms (the time the stimulus was displayed on the screen). A Wilcoxon signed rank test was used to

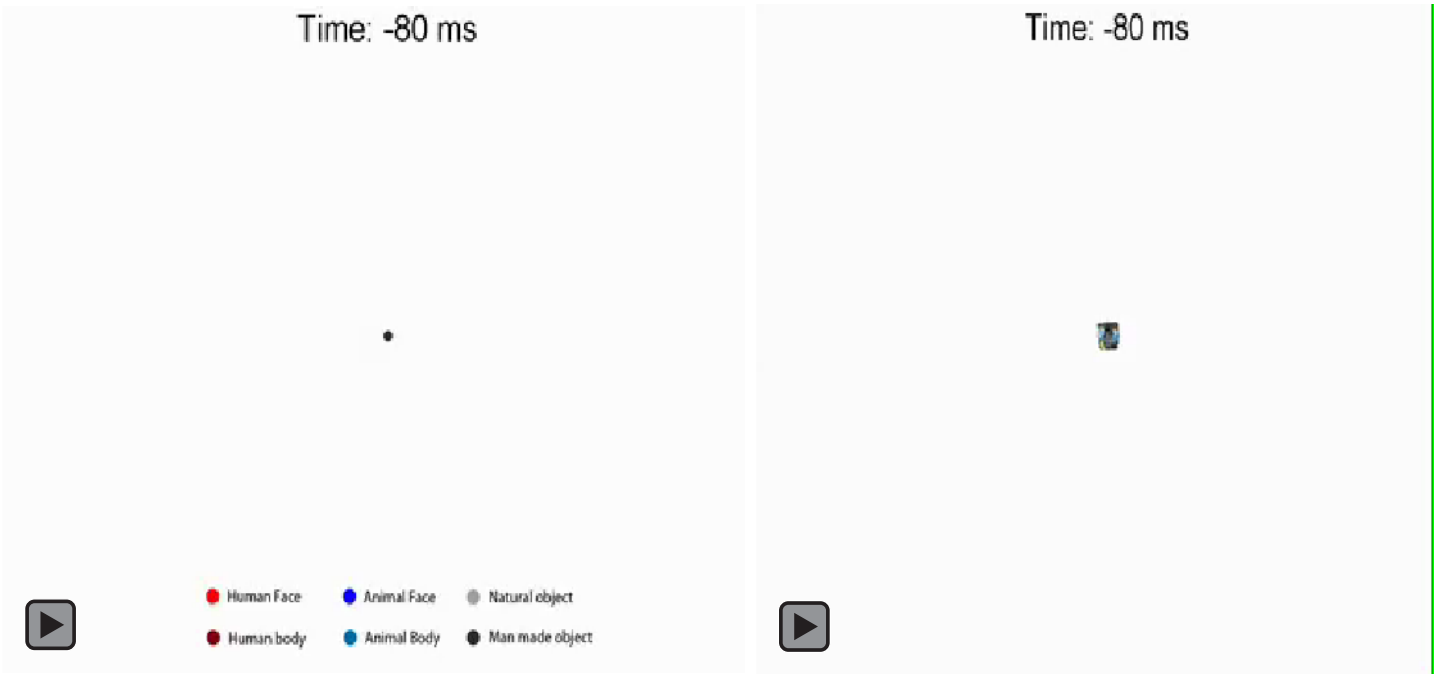


Figure 2. The emergence of object representations. The MDS movie graphically depicts how the brain's representation of the visual objects emerges over time. Distances between exemplar images represent the difference in brain activity between exemplars (i.e., decodability).

compare differences in peak latencies between levels in the stimulus hierarchy.

Temporal cross-decoding (TCD)

In the sliding-window decoding approach, classifiers are trained and tested on data from the same time point. This quantifies the information represented by the brain at particular times. How the informative patterns the enable decoding emerge over time can be studied by training classifiers at one time point and testing on other time points. The TCD procedure is to train classifiers for each time point in the time series; and then test each on the full time series. This approach has been used previously to study the dynamics of visual object processing (Carlson et al., 2011; see also Meyers, Freedman, Kreiman, Miller, & Poggio, 2008) and human perceptual decision making with EEG (Philiastides & Sajda, 2006).

For the TCD analysis, we opted to only examine the animate/inanimate object category decoding contrast, as this in one of the most prominent categorical distinctions in IT pattern response organization (Kiani et al., 2007; Kriegeskorte et al., 2008). We also chose to utilize IMCV because both approaches had qualitatively similar results and IMCV had higher decoding accuracy. To allows us to compare differences between brain activity associated with animate and inanimate objects across different time points, we modified the LDA classifier to use a single covariance estimate, which was estimated using all the of the time points. The procedure removes the influence of time varying changes in noise on the classifier. This procedure was also used in the analysis of scalp topographies, as this analysis similarly aimed to study changes in brain activity associated with animate and inanimate objects across different time points.

Results

Multidimensional scaling: The emergence of category divisions

We first examined the IMCV decoding data in an unsupervised fashion using MDS (Figures 2A and 2B). The distances between exemplars on each frame of the movie represent their dissimilarity in the brain's representation, i.e., decodability between exemplars. In using MDS, there is no presupposition of categorical structure. Nevertheless, categorical structure is prominent after the brain begins to process the stimuli. In the early time points (<60 ms), the exemplars are very close to one another, i.e., they are relatively indistinguishable based on brain activity. This is expected because the stimuli have not yet been presented before 0 ms, and the arrangement thus reflects the noise. Between 0 and 60 ms visual inputs from the retina have yet to reach

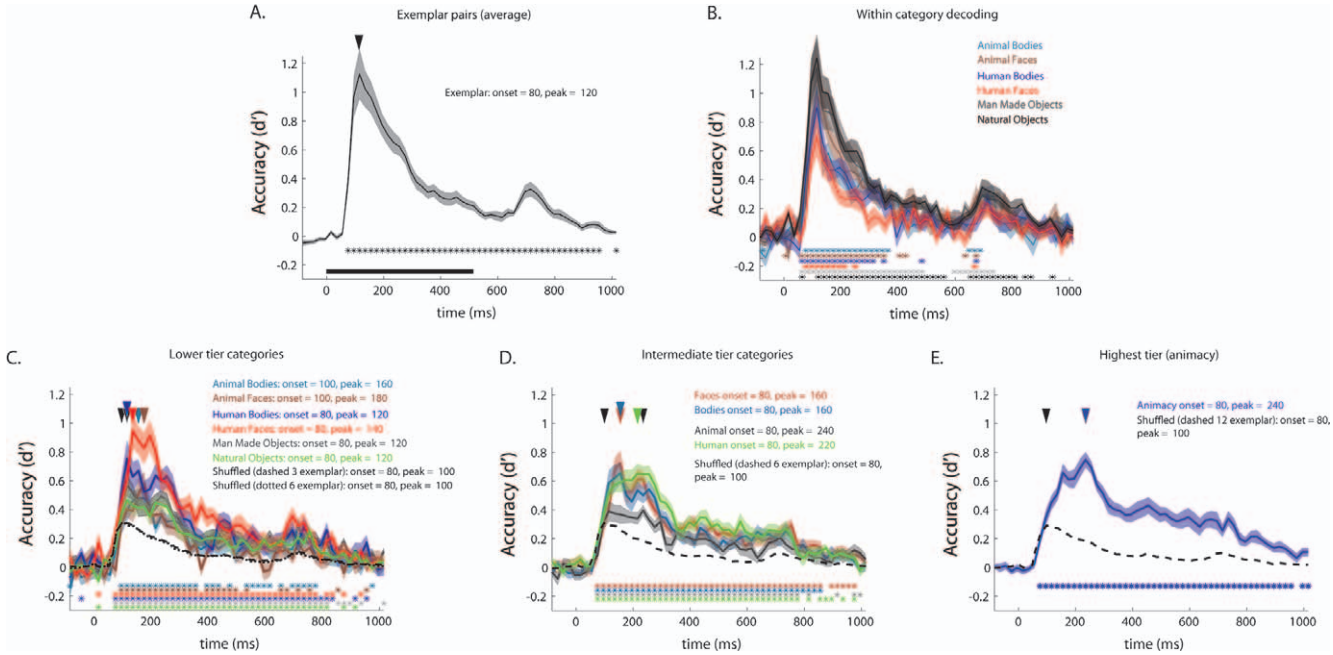


Figure 3. Emergence of exemplar and category discriminability for IMCV. (A) Average discriminability (d′) for all exemplar pairs. (B) Within category exemplar discriminability (d′) for Level 1 exemplar pairs. Exemplars are discriminable within each category. (C), (D), & (E) Discriminability of each Level 1, Level 2, and Level 3 category from stimuli outside the categories. The dashed line is average category decoding performance for 100 arbitrary categories (i.e., a categories comprised of randomly assigned exemplars). Solid lines are d′ averaged across subjects. The shaded region is 1 *SEM* across subjects. Color coded asterisks below the plots indicate above chance performance, as evaluated by a Wilcoxon signed rank test with a threshold of $p < 0.01$. Peak performance is indicated by color coded arrows above the plots. The onset and peak latencies are reported in the figure legends. The thick solid line below Plot A indicates the time the stimulus was on the screen (stimulus duration: 500 ms).

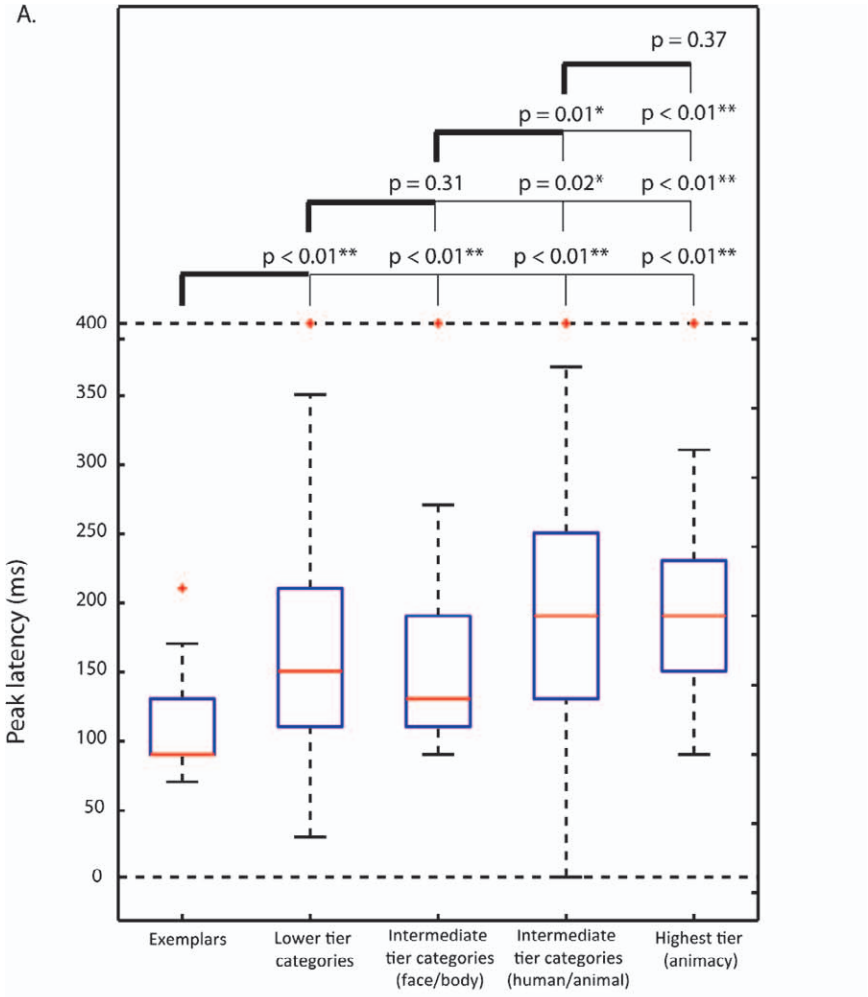
the cortex (Aine, Supek, & George, 1995; Brecelj, Kakigi, Koyama, & Hoshiyama, 1998; Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002; Jeffreys & Axford, 1972; Nakamura et al., 1997; Portin, Vanni, Virsu, & Hari, 1999; Supek et al., 1999), thus decoding exemplars is still at chance. The arrangement changes dramatically at 80 ms. At this point, there is a marked increase in the distance between exemplars, which suggests that individual exemplars are decodable (statistical inference below). The arrangement also suggests categorical structure (grouping of points representing exemplars of the same category), which becomes more prominent after 120 ms. Note how in the figures the human face stimuli cluster and distance themselves from the group, and the monkey face stays with this group until 180 ms. At this point, the human face stimuli cluster alone, which is consistent with a distinct response to human faces (Bentin et al., 1996; Liu et al., 2002). From 120 ms onward, object categories appear to cluster to varying degrees. The most notable distinction is that between animate and inanimate objects. After 160 ms, the animate and inanimate exemplars separate and remain distinguished. This is compatible with Kriegeskorte et al. (2008) and Kiani et al. (2007), who used a similar approach to discover categorical structure using MDS, and a range of studies showing differences in the representation of categories in the brain (Caramazza & Shelton, 1998; Chan, Halgren, Marinkovic, & Cash, 2011; Chao, Haxby, & Martin, 1999; Epstein & Kanwisher, 1998; Kanwisher et al., 1997; Konkle & Oliva, 2012; McCarthy, 1995; Shinkareva et al., 2008). In particular, the animate/inanimate dichotomy emerges as a prominent division (Caramazza & Mahon, 2003).

MDS provides a data-driven and descriptive global view of the results. We now describe the hypothesis-driven inferential analyses that address the questions posed at the outset.

The emergence of exemplar representations

We first studied the brain's individuation of the exemplars by examining the IMCV pairwise exemplars comparisons. Figure 3A shows the average accuracy across all possible pairwise combinations (264 comparisons total) and participants. In accordance with the MDS results, the exemplar decoding reached significance at 80 ms. The peak accuracy from exemplar

A.



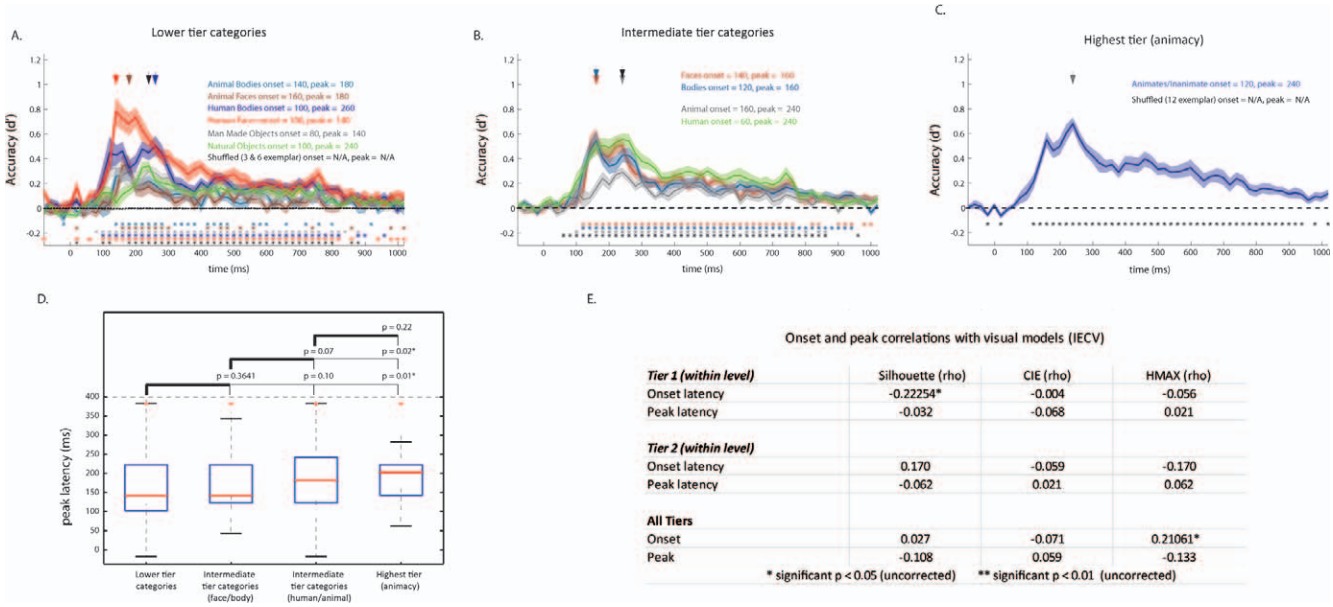
B.

Onset and peak correlations with visual models (IMCV)

Tier 1 (within level)	Silhouette (rho)	CIE (rho)	HMAX (rho)
Onset latency	-0.21749*	-0.138	-0.136
Peak latency	0.17664*	-0.200*	-0.099
Tier 2 (within level)			
Onset latency	0.204	-0.005	-0.165
Peak latency	0.065	0.076	-0.006
All Tiers			
Onset	-0.007	-0.033	-0.092
Peak	0.072	-0.102	0.061

* significant p < 0.05 (uncorrected) ** significant p < 0.01 (uncorrected)

Figure 4. Peak discriminability across categories for IMCV. (A) Decoding peak latencies for each level of the stimulus hierarchy (averages of within-category pairs). Central red lines indicate the median peak latency, box edges indicate the 25th and 75th percentiles, and whiskers indicate the most extreme values that are not outliers. Outliers are plotted as red crosses; those outside of 0 to 400 ms are plotted on the lower and upper bounds. Above the figure, the outcomes of Wilcoxon signed rank tests comparing levels of the stimulus hierarchy are summarized. Thick lines indicate the base comparison; thin lines indicate the comparison. A single asterisk indicates significance less than 0.05, double asterisks indicate significance less than 0.01. (B) Correlations category differences evaluated by visual models and observed onset and peak latency estimates. Asterisks indicate significant correlations (Spearman bootstrap test, uncorrected for multiple comparisons).



	Silhouette (rho)	CIE (rho)	HMAX (rho)
Tier 1 (within level)			
Onset latency	-0.22254*	-0.004	-0.056
Peak latency	-0.032	-0.068	0.021
Tier 2 (within level)			
Onset latency	0.170	-0.059	-0.170
Peak latency	-0.062	0.021	0.062
All Tiers			
Onset	0.027	-0.071	0.21061*
Peak	-0.108	0.059	-0.133

* significant $p < 0.05$ (uncorrected) ** significant $p < 0.01$ (uncorrected)

Figure 5. Category decoding with IECV. Panels A–C show decoding accuracy as a function of time for the three category levels. Here the decoder predicts whether a novel exemplar (not used in training) belongs inside or outside the indicated category. The dashed line is average category decoding performance for 100 arbitrary categories (i.e., a categories comprised of randomly assigned exemplars). Solid lines are d′ averaged across subjects. The shaded region is 1 *SEM* across subjects. Color coded asterisks below the plot indicate above chance performance, as evaluated by a Wilcoxon signed rank test with a threshold of $p < 0.01$. The onset and peak latency for each category is shown in the Figure legends. (A) Performance for lower tier category comparisons. (B) Performance for intermediate tier category comparisons. (C) Performance for the highest tier category comparison (animacy). (D) Summary boxplots for IECV Central red lines indicate the median peak latency, edges indicate the 25th and 75th percentiles, and whiskers indicate the extreme values that are not outliers. Outliers are shown as red crosses; those outside of 0 to 400 ms are plotted on the lower and upper bounds. Above the figure, the outcomes of Wilcoxon signed rank tests comparing levels of the stimulus hierarchy are summarized. Thick lines indicate the base comparison; thin lines indicate the comparison. Asterisks indicate significance less than 0.05. (E) Correlations category differences evaluated by visual models and observed onset and peak latency estimates. Asterisks indicate significant correlations (Spearman bootstrap test, uncorrected for multiple comparisons).

decoding was 100 ms. Thereafter, performance decays slowly. An apparent second peak can be seen at 740 ms, which is 220 ms after stimulus offset. In an earlier study, we similarly observed a second peak and showed it originated from the offset of the stimulus (Carlson et al., 2011). Since the second peak in the present study has a rough temporal correspondence to the offset, it presumably also reflects brain activity in response to the offset of the stimulus.

Decoding object exemplars within categories

Object exemplars are distinct in terms of their category (e.g., animate vs. inanimate) and low-level visual features (e.g., edges, color, etc.). Decoding performance therefore might reflect the brain's representation of an exemplar's category and/or the visual features associated with a specific exemplar. To study the brain's representation of features that distinguish exemplars within a category of objects, we analyzed exemplar discriminability within each category by delineating pairwise exemplars comparisons within categories. The average decoding of within-category exemplar pairs for each categories is shown in Figure 3B. The results show that exemplars within categories are decodable. Given the short onset and peak latencies, within-category exemplar discriminability is likely based on feature representations in early visual areas.

Category decoding without generalization to novel exemplars

We next examined how category structure emerges using the categorical comparisons from IMCV. Our stimuli had a planned hierarchical organization with three tiers (low, intermediate, and high) corresponding to the level of abstraction. Each category was represented by at least three exemplars, which vary in terms of their visual features (see Visual model analysis in Methods). Category decoding therefore

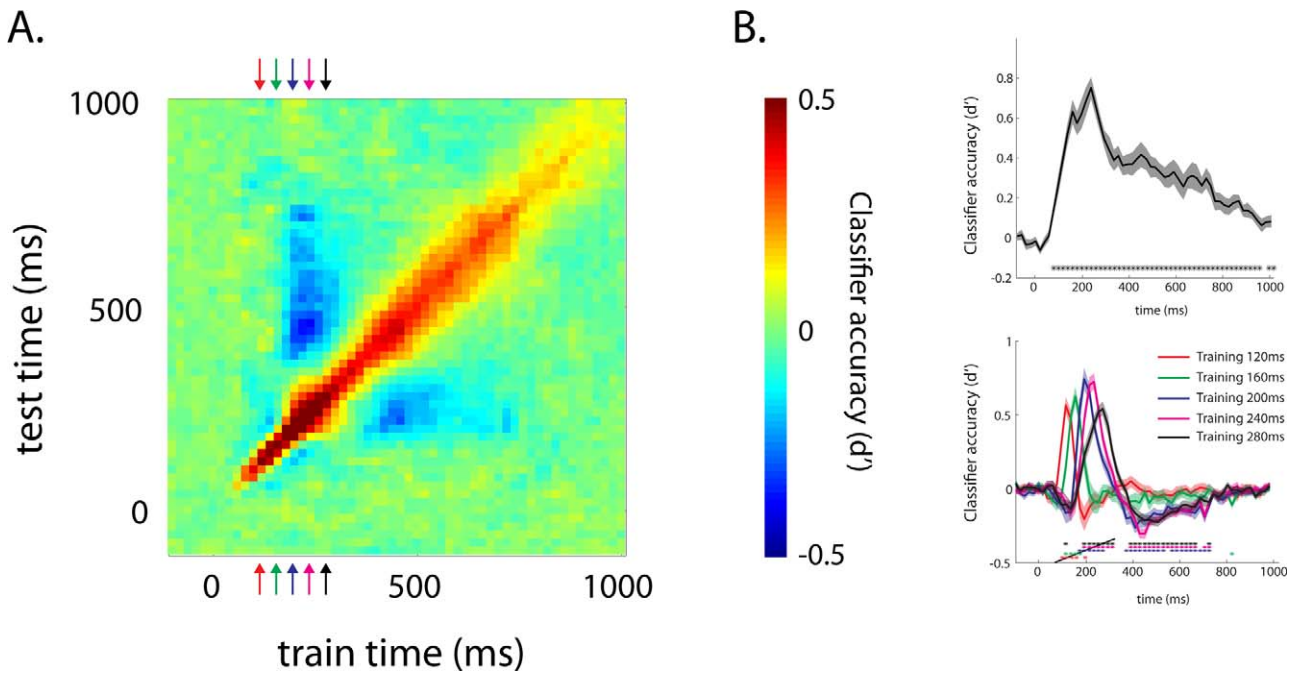


Figure 6. Discriminant cross training (A) Columns in the image are the time points the classifier was trained; rows are the times the classifier was tested. Color values indicate accuracy. Color-coded arrows above and below the image denote the times that the five classifiers in lower panel of B were trained. (B) The upper panel of the figure shows accuracy across subjects when the classifier is trained and tested on the same time point (i.e., sliding time window analysis). The shaded region indicates ±1 *SEM*. Black asterisks below the plot indicate accuracy significantly above or below chance. The lower panel shows accuracy for four classifiers trained at 100 ms, 140 ms, 180 ms, 220 ms, and 260 ms. Color-coded asterisks indicate accuracy significantly above or below chance. The line drawn through the asterisks indicates the times that training and test occurred on the same time point.

must rely on the brain's representation of features shared across the exemplars within a category, which are indicative of a category, and/or an explicit category representation. Figures 3C–E show decoding accuracy as a function of time indicating discriminability of the exemplars of a given category from the exemplars outside that category. The hierarchical category tiers are shown in separate plots with each category shown as a separate trace. Decoding for categories created by randomly assigning exemplars to categories with the same number exemplars is shown in the figure as a dashed line. Decoding performance for each "shuffled" category of varying size is based on 100 randomly constructed categories. Across the shuffled categories of varying size, there was no variation in the onset and peak latency. Differences in latencies between categories thus are unlikely to be attributed to differences in the number of exemplars used in the comparisons.

Onset latency is the earliest time that one category of exemplars can be distinguished from other category exemplars. The onset latencies were early and stable (ranging from 80 to 100 ms) across categories. These times are comparable with exemplar individuation and similar to the observed onset for artificial categories (80 ms). The analysis shows that there is sufficient information to reliably decode object category information, including artificial categories, shortly after stimulus onset using IMCV.

Peak latency is the optimal time to distinguish a category. In the analysis of peak latency, we observed variations between categories and a relationship between peak latency and the level of abstraction of the category. The peak latencies for the lower tier categories ranged from 120 to 180 ms. Intermediate tier categories ranged from 160–240 ms. In the intermediate tier, there was an apparent distinction between the different groupings. When the lower tier categories were grouped as faces and bodies, peak decoding was 160 ms. In contrast, when the categories were grouped as human and animal, peak decoding was much later (240 ms and 220 ms, respectively). The highest tier (animacy) had a peak latency of 240 ms.

While there are variations in the peak latencies for categories within levels, there was a pattern in the data indicating higher order categories emerge later. We investigated this possibility first by conducting an analysis of variance (ANOVA) to test for differences between levels. In accordance with our early observations, we found a significant difference between the

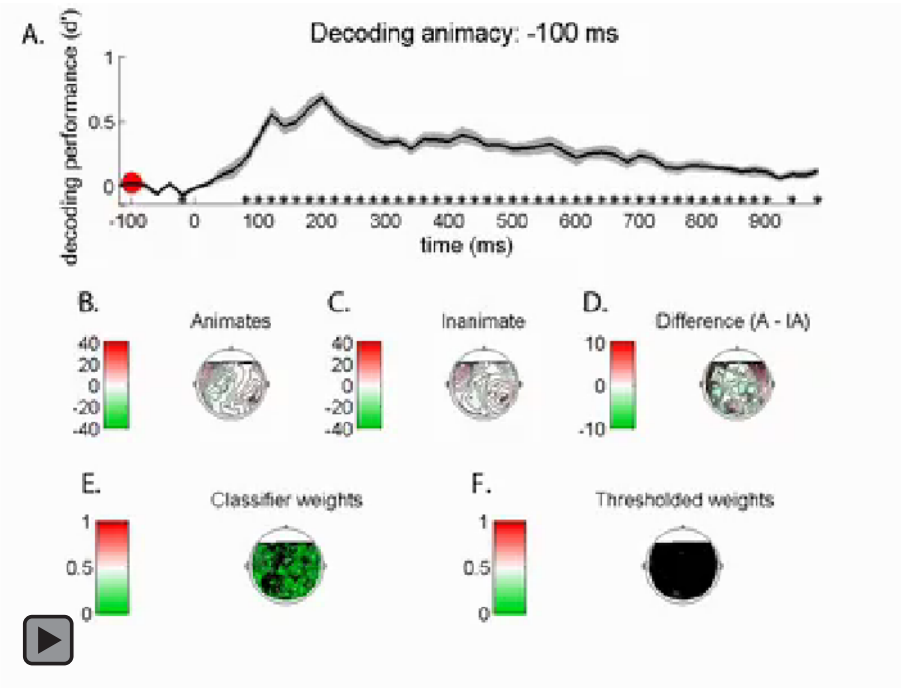


Figure 7. Scalp topographies for animate and inanimate objects and classifier weights (A) Average performance for animate/inanimate decoding. Solid lines are mean performance. The shaded region is 1 *SEM* across participants. Color coded asterisks below the plot indicate above chance performance, as evaluated by a Wilcoxon signed rank test with a threshold of $p < 0.01$. The superimposed transparent red denotes the time corresponding to the scalp topography and classifier weights shown in Plots B–F. (B)–(C) Average evoked scalp topography for animate and inanimate objects. (D) The difference in the evoked response between animate (A) and inanimate objects (IA). (E) Average absolute value of classifier weights across subjects. Data normalized to a range of 0–1 over the time series. (F) Thresholded classifier weights. Data from Plot D with sensors below statistical significance ($p > 0.05$) masked.

levels, $F(3, 265) = 5.4$; $p < 0.001$. We next tested for latency differences between individual tiers. The intermediate tiers of faces and bodies (Grouping 1) and human and animal (Grouping 2) were evaluated as separate in the analysis, as there was an apparent difference between these intermediate tiers in peak latency. Figure 4A summarizes the peak latency data with the outcomes of the statistical tests shown above the plot. The exemplar level was significantly earlier than all of the category levels ($p < 0.01$ for all comparisons). The category tiers sorted into two groups. Lower tier categories and the faces/bodies intermediate tier categories (Group 1) peaked first, and the comparison between these tiers was not significantly different. The human/animal intermediate tier categories (Group 2) and the highest tier (animacy) peaked later, and the comparison between these two tiers also was not significant. Between these divisions (lower tier and intermediate faces/bodies) and (intermediate human/animal and the highest tier), all of the comparisons were significant ($p < 0.05$).

In our analysis of feature differences between stimuli representing the categories using the silhouette, CIE-color, and HMAX models, we observed different models could distinguish one or two categories of stimuli. To study if feature differences could account for the onset and peak latency findings, we examined the relationship between each models capacity to distinguish a category and the onset and peak latency. Specifically, we tested if model correlations for category contrasts (i.e., the effectiveness of the model for a particular category) had a relationship to onset and peak decoding. The table in Figure 4B shows the correlation between model effectiveness and the observed onset and peak decoding performance. We found significant correlations (Spearman, bootstrap test) between the silhouette model and onset latencies within Tier 1 categories. This indicates that larger feature differences between Tier 1 categories result in shorter onset latencies. We also observed significant correlations between both the silhouette and CIE models and peak latency within the Tier 1 categories. Here, the interpretation is less clear because the two models differ in their sign (one positive, one negative). Thus in one case larger differences result in shorter peak latencies, and in the other larger differences result in longer peak latencies. Without a clear interpretation, we tentatively ascribe these correlations to randomness and the use of a liberal threshold ($p < 0.05$, uncorrected).

In summary, using IMCV we found that onset latencies were stable across categories, and that the

time object categories/exemplars are maximally distinct (i.e., peak decoding) depends on category and the tier of category abstraction. These findings suggest the brain may construct object category representations hierarchically.

Category decoding with generalization to novel exemplars

Individual exemplars have features that are both indicative (e.g., faces have eyes) and features that are not representative of their category (e.g., not all faces have green eyes). In the standard approach to decoding, cross validation is performed by dividing the data using some criterion, for example odd and even runs, to train and test the classifier (i.e., IMCV). Inherent in this procedure, data from individual stimuli (object exemplars) is included in both training and testing the classifier. As such, the classifier could learn to decode the category from brain activation patterns associated with specific features of exemplars. Standard cross validation is sensitive to representations specific to exemplars and also abstract category information. This is exemplified by the artificial categories in the IMCV. Despite the absence of a true category, the classifier could still learn to decode exemplars within the artificial category from exemplars outside the artificial category. In a separate analysis, we studied this further by cross validating using novel exemplars to test the classifier (IECV, see Methods). IECV is more stringent than standard cross validation in that decoding must rely on brain activation patterns shared by exemplars within a category.

Figure 5 summarizes the IECV decoding results. Panels A–C show the time series data. Artificial categories with the same number of exemplars are shown as dashed and dotted lines. For the artificial categories, the classifier was at chance. This is expected as the classifier was tested with novel exemplars and there is no systematic relationship between the exemplars in the artificial category. In examining the category contrasts, the results from IECV and IMCV are qualitatively similar, e.g., the conspicuous double peak in the human body decoding is maintained across methods. In accordance with IECV being a more conservative approach, accuracy was generally lower for IECV than IMCV. The onset latencies were also more variable. Onset latencies ranged from 60 ms (humans, Level 2b) to 160 ms (animal faces) for IECV, in contrast to the narrow range of 80 to 100 ms for IMCV. The increase in onset variability is expected because the classifier can only use activation patterns shared across exemplars within a category. Novel-exemplar decoding will thus be more sensitive to the heterogeneity of features of exemplars within a category. This is supported by the human face and human body comparisons. These two categories could be distinguished from the other categories based on early visual features (see models outcome in Table 1), and qualitatively appear to be relatively homogenous in terms of their features. Accordingly, both categories have early onsets using IMCV and IECV. The categories of animal faces and animal bodies, in contrast, appear more heterogeneous and could not be distinguished by any of the visual models. Correspondingly, onset latencies for these categories shift from short latencies in IMCV to longer latencies for IECV.

Panel D shows boxplots summarizing the peak latency data. Although the statistical outcomes were weaker for IECV, there was agreement between the two cross validation methods in terms of peak latencies. Median estimates of peak latency were within 20 ms (one time point) for all the levels (Figure 5D). An ANOVA conducted to examine differences in latencies between levels showed a marginal effect, $F(2, 132) = 2.64$; $p = 0.0749$. In accordance with IMCV, lower tier categories peak earlier compared to the highest level, and the difference between lower tier categories and the intermediate face and body groupings was not significant. The comparison between lower tier categories and the intermediate human and animal was marginal ($p = 0.10$). If this marginal were taken as significant, this result would be in agreement with IMCV. The intermediate tier face and body groupings had shorter latencies than the highest tier ($p = 0.02$), and were marginally shorter than the intermediate tier human and animal groupings ($p = 0.07$). Again, if the marginal were taken significant, this is compatible with the findings using IMCV.

We analyzed the relationship between the output of the visual models and onset and peak latencies estimated using IECV. The table in Figure 5E shows the correlation between model effectiveness and the observed onset and peak decoding latencies. Again, we found a correlation between the onset within the Tier 1 categories and the silhouette model, which suggests that shape similarity within the Tier 1 categories can account for the decoding onsets. We also observed a significant correlation between the HMAX model and onset across all the categories in the hierarchy. The positive correlation, however, is difficult to interpret (larger feature differences result in longer latencies). This might also be attributed to our use of a liberal threshold. Again the peak latencies findings cannot be accounted for by any of the models.

In sum, IECV is a more powerful approach as the technique focuses on activation patterns shared across exemplars within a category. The method yielded lower performance overall and increased variability in onsets relative to same-exemplar decoding. Novel-exemplar decoding similarly evidenced that the brain hierarchi-

cally constructs object categories of increasing levels of abstraction, in agreement with the IMCV decoding results. It is notable that this approach is still limited by the experimental selection of the stimuli. If there are low-level feature differences shared within the experimentally selected set of exemplars (that are not necessarily representative of the category), then the activation patterns evoked by these features could be used by the classifier to decode the stimuli. The failure of the visual models to account for the observed latencies differences, however, moderates this possibility.

Representational dynamics in object coding

We next examined the dynamics of the neural information that underlies category decoding. The predominant categorical boundary observed by Kriegeskorte et al. (2008) and Kiani et al. (2007), and in our own study, was animacy. We therefore chose to focus our analysis on this distinction. TCD analysis shows how well classifiers trained at individual time points generalize to other time points. For example, a classifier is trained with data from 100 ms relative to stimulus onset and then is tested on all the time points in the time series. Decoding performance over the time series will reveal the times that the classifier (trained at 100 ms) can generalize, which gives an indication of the dynamics of the patterns carrying the information. Figure 6A shows the TCD results for decoding animacy using IMCV. The figure gives a synopsis of the dynamics of the brain's representation of animacy. Columns in the image represent the times that the classifiers were trained; rows represent the times that the classifiers were tested. In this representation, the data on the diagonal is equivalent to the sliding time window analysis (plotted in the upper panel of Figure 6B). In the lower panel of Figure 6B, the performance of classifiers trained at 120, 160, 200, 240 (the peak performance for the animacy comparison from sliding time window analysis), and 280 ms are plotted separately to exemplify different representational trajectories.

The classifier trained at 100 ms exhibits a relatively simple dynamic. Performance transiently rises in response to the onset of the stimulus and then returns to chance. Representations of the stimulus from 80 ms to approximately 140 ms are of this general nature (see Figure 6A). This cascading sequence of transient representations is concordant with a feed forward sweep of activity (VanRullen, 2007). The trajectory of the representation measured at 160 ms is more complex. Decodable brain activity rises in response to the stimulus. Afterward there is a brief period of time where accuracy falls below chance (i.e., the classifier is systematically guessing incorrectly). It is important to note that the representation contains information about the categorical distinction during periods that the classifier is below chance. This pattern of above and below chance performance shows that decodable information in the representation changes such that the patterns of activity that represents a category at some times are anticorrelated with patterns representing the same category at other times (see Carlson et al., 2011). Notably if at these times the classifier was trained and tested on the same time point, as in the sliding window analysis (see upper panel of Figure 6B), the classifier would be above chance because the weights and decision rule would be different. Decodable information is thus sustained over the times that the classifier is above and below chance, but the patterns of brain activity that enable decoding is changing. This dynamic is even more apparent in classifiers trained at 200, 240, and 280 ms. In each case, the representations are changing over extended intervals of time (from approximately 200 to 700 ms) with performance fluctuating above and below chance. One final noteworthy aspect of the data is the square region in the image in Figure 6A around 240 ms. The square region indicates that there is a period of time that the classifier generalizes over a larger interval of time (~220 to 300 ms), which is indicative of a period of relative stability. This can also be seen in the lower panel of Figure 6B. The classifiers trained at 240 and 280 ms show very similar trajectories. The peak performance of two classifiers is different, which is expected, as peak performance will follow the time the classifier was trained. Still, both classifiers rise above chance about the same time and invert to below chance performance following nearly identical trajectories. This indicates that both classifiers are relying on the same representational state to decode animacy. This period of stable representation is followed by a large period of *below* chance performance in the row (large blue swath above the square in Figure 6A). One interpretation of this is that the representation inverts its activation profile, possibly due to excitation followed by inhibition and/or adaptation. Interestingly, the time of the stable representation starts around the time of peak classification performance (240 ms, see Figure 3D), the optimal time for readout of category information. Finally it is notable the information decays while the stimulus is still on the display, suggesting that the representation is maintained only until the information is transmitted. This further illustrates the dominant role of internal dynamics, as opposed to a stable response to the stimulus.

We next examined the dynamics of object coding by studying the scalp topographies and classifier weights. The movie in Figure 7 shows decoding performance for animate and inanimate objects (Figure 7A), the average evoked scalp topography for the two object categories

(Figures 7B–C), the difference between the scalp topographies (Figure 7D), and average the classifier weights across participants (Figures 7E–F). As can be seen in Figures 7B–C, the evoked response to animate and inanimate objects change over time but very similar. The difference (Figure 7D) reveals that the differential brain activity changes over time. Early in the time series, the difference is centered over visual cortex, as one might expect. Beyond this time, the difference shift to include brain activation patterns over virtually the entire brain. Notably the difference reveals little stability over time indicating that the patterns of brain activity that differentiate the two categories of objects continually change over time. For virtually of all of the time points, the topography of the weights exhibits little spatial coherence (Figures 7E; see also figure 5 of Chan et al., 2012). There was also little consistency in the topography of the weights over time. For each time point, the weights *appear* to be unique with little resemblance to other time points. This accords with the outcome of the TCD analysis; individual classifiers trained at one time point generalize poorly to other time points. This supports the contention that the neural information that distinguishes animate and inanimate objects in the brain is dynamic. Finally, we examined whether there was any consistency in the weights across subjects (Figure 7F). To do so, the classifier weights were thresholded by statistical significance using a liberal threshold ($p < 0.05$, uncorrected). At each time point only a few of the locations in the sensor array survived threshold, and those that did appear disbursed randomly and change over time. This random dispersion presumably reflects the use of a liberal uncorrected statistical criterion. The data thus indicate the classifier weights idiosyncratic. While the spatial variability of the weights described above and the idiosyncrasies across individuals on the surface might appear uninformative, there is one important implication. Traditional analyses in which a cluster of sensors are selected for analysis, e.g., sensors covering visual cortex, may not be able to detect the subtle differences the scalp topography between stimulus conditions. Pattern analysis approaches in contrast can make use of these differences to successfully decode stimulus conditions.

In summary, our analysis of the representational dynamics of the coding of animacy revealed representational trajectories that are transient and sustained. Our findings support the idea that the brain may use transient codes that stabilize only in a brief time window to represent visual objects. Finally, our data also suggest that pattern analysis approaches can reveal decodable information in the brain that is not immediately apparent from traditional analysis and thus can be used compliment these traditional approaches.

Discussion

Our study provides a new perspective on visual object processing and categorical divisions by revealing the emergence and decay of the information over the first 1000 ms. We used 24 object exemplars with a planned category structure, following previous primate single-unit recording studies and human fMRI studies of pattern representations (Kiani et al., 2007; Kriegeskorte et al., 2008). These studies showed that human and primate IT represents visual objects with coarse structure that distinguishes categories of objects (e.g., animate and inanimate) and fine-grained structure that distinguishes exemplars within categories. By marrying this approach with MEG, we were able to reveal the temporal dynamics of object categorization in the human brain.

Our study used naturalistic stimuli that aimed to engage visual areas specialized for the processing of real-world objects. In order not to compromise naturalism, we did not control for low-level feature differences (e.g., contrast, shape, color). We investigated the impact of feature representations on decoding category by examining several visual models (silhouette, CIE color, and HMAX) and by employing a second decoding procedure, in which the classifier decoded novel exemplars. Our analysis of the visual models found these models at most could discriminate 2 of the 10 categories, were only modestly successful in accounting for onset latency, and failed to account for our observed differences in peak latency. Decoding using novel exemplars (IECV) largely mirrored the findings using IMCV.

Studies have found that IT neurons have shorter latencies for human faces than other object categories and shorter latencies for human faces than animal faces (Bell et al., 2011; Carmel & Bentin, 2002; Kiani et al., 2005; McCarthy, 1995). The implication of these studies is that category information relates to response latency. In the present study, we examined this relationship using two latency measures: onset latency and peak latency. Using a standard decoding approach (same-exemplar decoding), we found little variation in onset latency by category. Decoding exemplars and categories narrowly ranged from 80–100 ms. Several of our findings indicate that early decoding (<100 ms) is feature based. The raw timing is near the time that visual information first reaches the cortex (Aine et al., 1995; Brecelj et al., 1998; Di Russo et al., 2002; Jeffreys & Axford, 1972; Nakamura et al., 1997; Portin et al., 1999; Supek et al., 1999). We observed greater variation in onset latencies decoding novel exemplars, a procedure that controls for low-level feature differences between stimuli. We found similar onset latencies (80–100 ms) for decoding exemplars within categories and decoding exemplars based on artificial categories.

Finally, our previous work has found that showing stimuli in varied spatial locations, which acts as a control for low-level feature differences, delays onset latencies to 110 ms or greater (Carlson et al., 2011). Although early visual representations may not be explicit categorization by the brain per se, behavioral studies have shown that humans use information represented in early visual areas to categorize objects (Honey, Kirchner, & VanRullen, 2008). Categorization in the brain thus might be better viewed as a process of accumulating evidence for category membership. If so, our findings suggest that the brain could read out a variety of object categories as fast as 80 ms based on biases in early feature representations. This might account for some of the remarkable behavioral findings in rapid categorization (Honey et al., 2008; Kirchner & Thorpe, 2006). As a methodological aside, decoding of objects from feature representations including highly homogenous sets of object within a category (e.g., faces, see Figure 3B), demonstrates the sensitivity of MEG decoding methods (and presumably also EEG), which suggests that model based decoding approaches (Dumoulin & Wandell, 2008; Kay, Naselaris, Prenger, & Gallant, 2008) for MEG/EEG may be a promising avenue for future research.

Recent studies have suggested perceptual categorization is process of accumulating evidence that depends on the strength of the perceptual evidence by showing that neural activity that correlates with behavioral performance was delayed with increasing task difficulty (Philiastides & Sajda, 2006). If categorization is viewed as a process of accumulating evidence, the peak latency would be the optimal time to "read out" category information from neuronal activation patterns. While we observed differences between same-exemplar decoding and novel-exemplar decoding in terms of onset latency, we found that the two approaches were in good agreement for the peak latency measure, excluding the exemplar level, which was not included in the novel-exemplar decoding analysis. In a comparison of the two methods by individual categories, only two of the eleven categories (human bodies and natural objects) differed by more than 20 ms (one time point) in peak latency estimates. In our analysis of the categorical tiers of the hierarchy, the two methods showed good agreement (all estimates of peak latency were within 20 ms). We interpret the cohesion of the two methods on peak latency estimates to indicate that feature differences between exemplars do not have a great impact on peak latency.

For peak latency, we observed category structure emerges in accordance with our planned hierarchy. Our findings thus show that the brain delineates lower tier object categories first along with the intermediate face and body categories, and then higher order categories emerge. We observed one notable discrepancy from the hierarchical stimulus structure. In the intermediate tier, peak latencies for faces and bodies coincided with lower tier categories. This discrepancy might be reconciled by the conjecture that the brain has special mechanisms to process faces and bodies due to their ecological significance and/or our extensive perceptual experience with these categories, a conjecture supported by a large body of research, which has shown there are areas in human and primate ventral temporal cortex that respond preferentially to faces and bodies (Bell et al., 2011; Downing et al., 2001; Kanwisher et al., 1997) and neurons in IT that are selective for faces and body parts (Bell et al., 2011; Desimone, 1991).

In interpreting our data, we do not imply that the brain categorizes objects in stages that correspond to the experimentally defined tiers in our hierarchical stimulus structure, which were arbitrarily defined. Instead, we take the broad perspective that the categorization of objects by the brain is a process of accumulating evidence (Philiastides & Sajda, 2006). In this view, the finding that category structure emerges from specific to abstract has the implication that category representations resolved early can provide supporting evidence to category representations that are constructed later, e.g., activation of the category "human face" would support the representations "human" and "animate." We also would not advocate the strong conclusion that object recognition takes places in a strict hierarchical fashion from specific to abstract. Nonhuman primates studies have shown that responses to categories (e.g., face) emerge prior to subordinate members (Matsumoto et al., 2005; Sugase et al., 1999); a finding that directly contradicts this interpretation. In reconciling these findings, one could posit a plausible entry point to be basic level categories (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), which humans act faster to categorize than more abstract and specific instances of categories. Basic level categories as an entry point for object recognition could be advantageous as the identification of stimuli's basic level category could constrain subordinate categories that the stimulus belong to and facilitate identification. This basic level category entry point hypothesis could be explicitly tested in future research.

In recent years, research has begun to acknowledge the importance of temporal dynamics in population codes (Buonomano & Maass, 2009; Rabinovich et al., 2008; Stokes, 2011). In particular, populations of neurons may encode information through systematic changes in the response patterns over time. We used TCD to study how population responses, measured using MEG, elicited by visual objects change over time. The TCD approach is similar to the notion of a *virtual electrode* except it centers on a particular representational state, as opposed a source of activity in the brain. Classifiers are used to capture and characterize

information in a specific representational state, and the dynamics of the information in this state is evaluated as a function of time. We found that exemplar and category information can be decoded from brain activity from approximately 80–800 ms using a sliding time window. In our analysis of the representational dynamics, we found representational states that discriminate category membership can be transient or sustained. Our analysis centered on the coding of animacy (Figure 6), although the other object categories exhibited similar dynamics (data not shown). In the coding of animacy, we found early representations (60–120 ms) are transient. In the range of 220 to 300 ms, brain representations were sustained; however, the representation of category information changed over time. Empirical and modeling studies have argued dynamic population codes can confer sensory representations the capacity to encode time (Buonomano & Maass, 2009), memories of past sensory events (Nikolic et al., 2009), and behavioral goals (Crowe et al., 2010; Woloszyn & Sheinberg, 2009). Whether the brain uses dynamic codes to represent information relevant to object perception is an open question. Furthermore, it is useful to consider more basic questions like whether the brain "reads out" category information from these time varying patterns (cf. Williams et al., 2007), and if so how. If the brain does use temporal codes, at what time point does the brain begin to read out the signal and over what interval of time? It is also possible the brain does not use a temporal code for encoding object category information. The brain may instead simply read out category information at the optimal time (i.e., peak decodability) or the earliest time point that a decision can be made reliably (onset decodability).

Our findings can be summarized as follows: (1) Visual object categories can be distinguished by early brain representations encoding visual features, and biases in category features might be used by the brain to rapidly categorize stimuli; (2) the time the brain takes to maximally distinguishes categories (i.e., peak latency) is dependent on the level of category abstraction; and (3) the brain encodes category information using transient representations that dynamically change over time. Collectively, these findings elucidate how visual objects are represented in the brain. The brain initially represents visual objects in a cascading sequence of representations with limited lifetimes, concordant with a rapid feed forward sweep (VanRullen, 2007), In this cascade of activity, object representations emerge. The sequencing of object representations at different levels is advantageous, as the immediate categorization of an object at one level, possibly at the so called basic level (Rosch et al., 1976), can inform more abstract representations and constrain the subset of candidate subordinate categories and identities of a stimulus.

Keywords: object recognition, categorization, magnetoencephalography, pattern-information analysis, brain decoding

Acknowledgments

This work was funded by an Australian Research Council Future Fellowship to T. C. (FT120100816), and in part by the Medical Research Council of the UK (programme MC-A060-5PR20) and by a European Research Council Starting Grant (ERC-2010-StG 261352) to N. K. We would also like to thank the reviews for their helpful suggestions.

Commercial relationships: none.
Corresponding author: Thomas A. Carlson.
Email: thomas.carlson@mq.edu.au.
Address: Department of Cognitive Sciences, Centre for Cognition & its Disorders, Macquarie University, Sydney, NSW, Australia.

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.

Bell, A. H., Malecek, N. J., Morin, E. L., Hadj-Bouziane, F., Tootell, R. B., & Ungerleider, L. G. (2011). Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity. *The Journal of Neuroscience, 31,* 12229–12240.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face prception in humans. *Journal of Cognitive Neuroscience, 8,* 551–565.

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.

Buonomano, D. V., & Maass, W. (2009). State-dependent computations: Spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience, 10,* 113–125.

Caramazza, A., & Mahon, B. Z. (2003). The organization of conceptual knowledge: The evidence from category-specific semantic deficits. *Trends in Cognitive Sciences, 7,* 354–361.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the animate-inanimate distinction. *Journal of Cognitive Neuroscience, 10,* 1–34.

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*(10):9, 1–17, http://www.journalofvision.org/content/11/10/9, doi:10.1167/11.10.9. [PubMed] [Article]

Carmel, D., & Bentin, S. (2002). Domain specificity versus expertise: Factors influencing distinct processing of faces. *Cognition, 83,* 1–29.

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.

Chan, A. M., Halgren, E., Marinkovic, K., & Cash, S. S. (2012). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *Neuroimage, 54*(4), 3028–3039.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience, 2,* 913–919.

Crowe, D. A., Averbeck, B. B., & Chafee, M. V. (2010). Rapid sequences of population activity patterns dynamically encode task-critical spatial information in parietal cortex. *The Journal of Neuroscience, 30,* 11640–11653.

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

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience, 3,* 1–7.

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. (2001). A 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* (pp. xx, 654). New York, NY: Wiley.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage, 39,* 647–660.

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

Haxby, J. V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430.

Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid 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.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura [Translation: Distribution of alpine flora in the basin Dranses River and some neighboring areas]. *Bulletin de la Société Vaudoise des Sciences Naturelles, 37,* 547–579.

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.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8,* 679–685.

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

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature, 452,* 352–355.

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology, 97,* 4296–4309.

Kiani, R., Esteky, H., & Tanaka, K. (2005). Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *Journal of Neurophysiology, 94,* 1587–1596.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46,* 1762–1776.

Konkle, T., & Oliva, A. (2012). A real-world size

organization of object responses in occipitotemporal cortex. *Neuron, 74,* 1114–1124.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron, 60,* 1126–1141.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience, 5,* 910–916.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19,* 577–621.

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 U S A, 92,* 8135–8139.

Matsumoto, N., Okada, M., Sugase-Miyamoto, Y., Yamane, S., & Kawano, K. (2005). Population dynamics of face-responsive neurons in the inferior temporal cortex. *Cerebral Cortex, 15,* 1103–1112.

Mazor, O., & Laurent, G. (1995). Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron, 48,* 661–673.

McCarthy, R. A. (1995). *Semantic knowledge and semantic representations*. New York, NY: Psychology Press.

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology, 100,* 1407–1419.

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.

Nikolic, D., Hausler, S., Singer, W., & Maass, W. (2009). Distributed fading memory for stimulus properties in the primary visual cortex. *PLoS Biology, 7,* e1000260.

Philiastides, M. G., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex, 16,* 509–518.

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.

Rabinovich, M., Huerta, R., & Laurent, G. (2008). Transient dynamics for neural processing. *Science, 321,* 48–50.

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

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

Shinkareva, S. V., Mason, R. A., Malave, V. L., Wang, W., Mitchell, T. M., & Just, M. A. (2008). Using FMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS One, 3,* e1394.

Stokes, M. (2011). The spatiotemporal structure of population coding in monkey parietal cortex. *The Journal of Neuroscience, 31*(4), 1167–1169.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature, 400,* 869–873.

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.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York, NY: Wiley.

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

VanRullen, R. (2007). The power of the feed-forward sweep. *Advances in Cognitive Psychology, 3,* 167–176.

VanRullen, R. (2011). Four conceptual fallacies in mapping the time course of recognition. *Frontiers Perceptual Psychology, 2,* 365.

Williams, M. A., Dang, S., & Kanwisher, N. G. (2007). Only some spatial patterns of fMRI response are read out in task performance. *Nature Neuroscience, 10,* 685–686.

Woloszyn, L., & Sheinberg, D. L. (2009). Neural dynamics in inferior temporal cortex during a visual working memory task. *The Journal of Neuroscience, 29,* 5494–5507.