Feature-coding transitions to conjunction-coding with progression through human visual cortex

Rosemary A. Cowell[1,*]

Krystal Leger[1]

John T. Serences[2,3]

[1]Department of Psychological and Brain Sciences, University of Massachusetts, Amherst, MA 01003, USA

[2]Department of Psychology, University of California, San Diego, La Jolla, CA 92093-0109, USA

[3]Neurosciences Graduate Program, University of California, San Diego, La Jolla, CA 92093, USA

*Corresponding author – Contact Information:

Email: rcowell@umass.edu

Tel: +1-413-687-7855

Address:
Department of Psychological and Brain Sciences
University of Massachusetts
135 Hicks Way
Amherst, MA 01003
USA

Number of Pages: 57; Number of Figures: 9; Number of Tables: 7; Abstract: 247 words; New and Noteworthy: 66 words

Keywords: visual cortex, feature, conjunction, neuroimaging, object representations

Author Contributions (CRediT taxonomy)

Conceptualization, R.A.C.; Methodology, R.A.C. and J.T.S; Software, R.A.C., K.L. and J.T.S.; Formal Analysis, R.A.C., K.L and J.T.S.; Resources, R.A.C. and J.T.S.; Writing – Original Draft, R.A.C.; Writing – Review and Editing, R.A.C. and J.T.S.; Visualization, R.A.C; Funding Acquisition, J.T.S and R.A.C.

Abstract

Identifying an object and distinguishing it from similar items depends upon the ability to perceive its component parts as conjoined into a cohesive whole, but the brain mechanisms underlying this ability remain elusive. The ventral visual processing pathway in primates is organized hierarchically: Neuronal responses in early stages are sensitive to the manipulation of simple visual features whereas neuronal responses in subsequent stages are tuned to increasingly complex stimulus attributes. It is widely assumed that feature-coding dominates in early visual cortex whereas later visual regions employ conjunction-coding in which object representations are different from the sum of their simple-feature parts. However, no study in humans has demonstrated that putative object-level codes in higher visual cortex cannot be accounted for by feature-coding and that putative feature-codes in regions prior to ventral temporal cortex are not equally well characterized as object-level codes. Thus the existence of a transition from feature- to conjunction-coding in human visual cortex remains unconfirmed, and, if a transition does occur, its location remains unknown. By employing multivariate analysis of functional imaging data, we measure both feature-coding and conjunction-coding directly, using the same set of visual stimuli, and pit them against each other to reveal the relative dominance of one versus the other throughout cortex. Our results reveal a transition from feature-coding in early visual cortex to conjunction-coding in both inferior temporal and posterior parietal cortices. This novel method enables the use of experimentally controlled stimulus features to investigate population-level feature- and conjunction-codes throughout human cortex.

New and Noteworthy

We use a novel analysis of neuroimaging data to assess representations throughout visual cortex, revealing a transition from feature-coding to conjunction-coding along both ventral and dorsal pathways. Occipital cortex contains more information about spatial frequency and contour than about conjunctions of those features, whereas inferotemporal and parietal cortices contain conjunction coding sites in which there is more information about the whole stimulus than its component parts.

Introduction

Object perception is underpinned by a hierarchical series of processing stages in the ventral visual pathway (Felleman & Van Essen, 1991; Hubel & Wiesel, 1965; Kobatake & Tanaka, 1994). At each successive stage from primary visual cortex (V1) to anterior inferotemporal (aIT) cortex, the complexity of the optimal stimuli increases: neurons in V1 are tuned to simple stimulus attributes such as orientation (Hubel & Wiesel, 1965; Mazer, Vinje, McDermott, Schiller, & Gallant, 2002); neurons in V4 and posterior inferotemporal cortex (pIT) are selective for moderately complex features (Brincat & Connor, 2004; Kobatake & Tanaka, 1994; Pasupathy & Connor, 1999; Rust & Dicarlo, 2010); and neurons in aIT prefer partial or complete views of complex objects (Desimone, Albright, Gross, & Bruce, 1984; Kobatake & Tanaka, 1994; Tanaka, 1996). Data from functional magnetic resonance imaging (fMRI) in humans corroborate these findings: the blood oxygenation level-dependent (BOLD) signal exhibits selectivity for orientation, spatial frequency and color in early visual regions (Brouwer & Heeger, 2009; Henriksson, Nurminen, Hyvarinen, & Vanni, 2008; Kamitani & Tong, 2005; Serences, Saproo, Scolari, Ho, & Muftuler, 2009), but is sensitive to object-level properties such as global contour or object category in higher visual regions (Drucker & Aguirre, 2009; Kanwisher, McDermott, & Chun, 1997; Kriegeskorte et al., 2008; Malach et al., 1995; Ostwald, Lam, Li, & Kourtzi, 2008). It is widely assumed that downstream, object-specific representations are constructed through combination of the simple feature representations upstream, but the manner in which this combination occurs remains unknown.

There are at least three possible combination schemes. The first assumes that downstream object-level representations perform ‘and-like’ operations on upstream feature representations (Rust & DiCarlo, 2012), transforming the feature-code into conjunction-sensitive representations in inferotemporal (IT) cortex. This *feature-to-conjunction transition* scheme is assumed by many models of object processing (Bussey & Saksida, 2002; Cowell, Bussey, & Saksida, 2010; Fukushima, 1980; Perrett & Oram, 1993; Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007; Wallis & Rolls, 1997) and accords with electrophysiological findings of IT neurons selective for complex objects (Desimone et al., 1984; Fujita, Tanaka, Ito, & Cheng, 1992; Kobatake & Tanaka, 1994). However, when tested with large stimulus sets many IT neurons show broad tuning, responding to multiple complex objects (Desimone et al., 1984; Kreiman et al., 2006; Zoccolan, Kouh, Poggio, & DiCarlo, 2007). Therefore, apparent object-level selectivity in an IT neuron tested with a small stimulus set might instead reflect selectivity for a low-level feature possessed by only a few objects in the set. Thus the data do not rule out a second scheme: a *global feature-coding hypothesis,* in which simple features are coded separately in visual cortex and are bound by the synchronization of neural activity rather than by convergence onto a cortically localized representation of the conjunction (Eckhorn, 1999a, 1999b; Singer & Gray, 1995). Finally, a third possible coding scheme is a *global conjunction-coding hypothesis*, in which all stations in the hierarchy bind simple features together non-linearly to produce conjunctions (Shigihara & Zeki, 2013, 2014). Under this scheme, the apparent feature-selectivity of early visual cortex belies a neural code that is optimized for discriminating complex objects that contain those features. Supporting this account, several studies have reported coding of simple conjunctions of features such as color, form, motion and orientation, in early visual regions (Anzai, Peng, & Van Essen, 2007; Engel, 2005; Gegenfurtner, Kiper, & Levitt, 1997; Johnson, Hawken, & Shapley, 2008; Seymour, Clifford, Logothetis, & Bartels, 2009, 2010), and coding of complex conjunctions in both early *and* higher-level visual regions (Erez, Cusack, Kendall, & Barense, 2015).

To differentiate between the three alternative schemes, we must measure not just the presence of feature-coding and conjunction-coding throughout visual cortex, but the *relative* contribution of each. Using fMRI in humans, we devised a novel stimulus set and multi-variate pattern analysis (MVPA) technique to pit feature-coding against conjunction-coding. We created stimuli by building conjunctions from binary features, thereby allowing each cortical region to be probed for information at two levels: features or conjunctions. Feature-information and conjunction-information were assessed using the same neuroimaging dataset and placed in a ratio, allowing direct comparison of the two coding schemes throughout cortex.

Methods and Materials

Methods Overview

Participants in the scanner viewed visual stimuli constructed hierarchically from four binary features to give sixteen unique, conjunctive objects. We verified by means of a one-back repetition detection task that participants were attending to the stimuli sufficiently well to discriminate between distinct objects, i.e., between different unique conjunctions of features (Table 1). We used a support vector machine (SVM) to classify patterns of BOLD responses evoked by the stimuli. For each session in each subject, we constructed four two-way feature-level SVM classifiers (one classifier for each binary feature) and one sixteen-way object-level SVM classifier (Figure 1). This yielded both feature- and object-level classification accuracy for a given region of interest (ROI) (Tables 2 and 3). We next constructed a *Feature Conjunction Index* (FCI) for each ROI by comparing the output of the feature- and object-level classifiers (Figure 1; Table 4). A positive FCI indicates that the ROI contains more information about individual objects than is predicted from the information it contains about the separate object features, suggesting that its activation pattern is modulated by the presence or absence of specific objects rather than by individual features. A negative FCI indicates that the ROI contains more information about individual features than about whole objects, suggesting that voxel activations are primarily modulated by individual feature dimensions rather than by whole-object identity. These interpretations of FCI were confirmed via analyses of synthetic data.

Participants, Stimuli, Task and Data Acquisition

Participants

Eight healthy participants (4 female) with normal or corrected-to-normal vision completed two scan sessions. All participants provided written informed consent as approved by the Institutional Review Board at UCSD, and were compensated at $20/hr for fMRI scan sessions and $10/hr for behavioral test sessions.

Stimulus and Task Parameters

Motivated by evidence that the integration of contour elements into global shape (Brincat & Connor, 2004) and local image features into global texture (Goda, Tachibana, Okazawa, & Komatsu, 2014; Hiramatsu, Goda, & Komatsu, 2011) are key mechanisms by which the ventral pathway constructs complex object representations, we created novel object stimuli by building conjunctions from binary features defined by contour and spatial frequency (Figure 1). To examine whether conjunction-coding emerged from an upstream feature-code, it was important to choose stimulus features that are encoded by early visual regions. Although shape contour is often considered a relatively high-level property of a visual stimulus and is known to be represented in areas like LOC, the binary contour features we used must be encoded as a collection of simple, oriented line segments in early visual regions such as V1 and V2, because of small receptive field size (see also Brincat & Connor, 2006; Yau, Pasupathy, Brincat, & Connor, 2013). The sensitivity of neurons in early visual cortex to spatial frequency is well-documented (e.g., Foster, Gaska, Nagler, & Pollen, 1985; von der Heydt, Peterhans, & Dursteler, 1992). Stimuli subtended approximately 7-10^0 of visual angle, except in one session in one participant (subject AF) in which they subtended approximately 5-7^0 (visual inspection of multi-variate pattern analysis (MVPA) results did not give any indication of greater between-session differences in any of the MVPA measures for subject AF than for other subjects). Visual displays were presented to participants via back-projection onto a screen at the foot of the scanner bore, which was viewed in a mirror fixed to the head coil, over a distance of approximately 380cm. To ensure that all pixel locations emitted the same average luminance over the course of a trial (and therefore that all stimuli possessed the same average luminance), visual stimuli cycled continuously from a positive to a negative image, with every pixel oscillating from minimum to maximum luminance according to a temporal sine wave with frequency of 2Hz.

Each scan session contained 10 experimental task runs. Each task run lasted 274 seconds (44 trials lasting 6 s each, in addition to a 10 s post-task scanning window). Participants were instructed to fixate a circular, colored fixation point that appeared at the center of the screen throughout all trials. Each run comprised 32 'stimulus' trials, 2-4 'immediate repeat' trials, and 8-10 'null' trials. On stimulus and immediate repeat trials, the fixation point was red, and stimulus onset began 0.2-0.7 s after the start of the trial (with exact onset randomly jittered within that window); stimulus presentation lasted 3 s and was followed by a 2.3-2.8 s response window in which only the fixation point appeared, to give a total trial duration of 6 s. The variable cue onset time produced an inter-stimulus interval that varied between 2.5 and 3.5 s. The 32 stimulus trials comprised 2 pseudo-randomly ordered presentations of each of the 16 unique stimuli. To generate immediate repeat trials, a pseudo-randomly chosen stimulus was inserted into the sequence such that it created a repeat of the stimulus in the immediately preceding trial; functional data from these trials were removed from multi-variate analyses. On null trials, the fixation point changed from red to green and participants were required to press any button whenever they detected a slight dimming of the green fixation point, which could occur once or twice per null trial; a response was required within 1 s of each dimming event for a trial to be scored as correct. This task was designed to reduce the tendency for mind-wandering, known to affect baseline measures of BOLD (Stark & Squire, 2001). Accuracy on null trials was monitored to ensure participant wakefulness, and the degree of fixation point dimming was adjusted between runs to produce below ceiling performance such that attention was maintained. Participants performed a one-back repetition detection task, indicating by button press whether the stimulus was the same (button 2) or different (button 1) from that of the previous trial. Good performance on this task required wakefulness and attention to the stimuli.

All participants were familiarized with the stimuli and task in a brief practice session before the first scan.

Visual Search Training

All subjects completed several daily sessions of discrimination training on the set of sixteen stimuli, interposed between the two scan sessions. These behavioral training sessions were conducted outside of the scanner. The task was adapted from a visual search task used by Shiffrin and Lightfoot (Shiffrin & Lightfoot, 1997). On each trial, a target stimulus chosen pseudo-randomly from the set of 16 was displayed singly, in the center of the screen, for 3 sec. Immediately after stimulus offset a search array appeared, which contained between 1 and 8 stimuli located at 8 equally eccentric spatial locations (the assignment of stimuli to locations was random, with slots left empty when the search array comprised fewer than 8 stimuli). Participants were allowed up to 20 seconds to indicate by button press whether the target was present or absent in the display. Feedback in the form of a low or high auditory beep indicated whether the response was correct or incorrect, respectively. In each daily session, participants completed ten blocks of 32 trials, comprising two trials with each of the 16 stimuli serving as target. Accuracy and response time (RT) data were collected. This visual search task required that participants attend to the specific conjunction of visual features comprising the target, in order to discriminate the target from the distractor stimuli, which shared features with the target. In conjunctive visual search, which requires inspection of each stimulus in series, RTs are typically longer when a subject must search a display containing more stimuli; this produces a positive slope for the relationship between search display set size and RT (Treisman & Gelade, 1980). However, Shiffrin and Lightfoot (1997) showed that training on a conjunctive visual search task causes RT-set size slopes to decrease in magnitude, presumably as the conjunctions become unitized such that 'pop-out' occurs, obviating the need for serial search. Because our aim was that participants' representations of the stimulus conjunctions would become unitized, the dependent variable of interest was the RT-set size slope and how it changed across daily training sessions. Accordingly, training was terminated for each subject when the RT-set size slope appeared to be approaching an asymptotically low value (mean 11.1 days; range 7 - 15 days). Participants completed their second scan within 3 days of the last training session.

Acquisition of fMRI data

Each scan session lasted 2 hours and included ten experimental runs and two retinotopic localizer runs. We scanned participants on a 3T GE MR750 scanner at the UCSD Keck Center for Functional Magnetic Resonance Imaging, using a 32 channel head coil (Nova Medical, Wilmington, MA). Functional data were acquired using a gradient echo planar imaging (EPI) pulse sequence with TR = 2000 ms, TE = 30 ms, flip angle = 90°, voxel size 2 × 2 × 3 mm, ASSET factor=2, 19.2 × 19.2 cm field of view, 96 × 96 matrix size, 35 slices of 33 mm thickness with 0 mm spacing, slice stack obliquely-oriented passing through occipital, ventral temporal, inferior frontal and posterior parietal cortex. The oblique orientation (i.e. tilted downward at the front of the brain) ensured good coverage of ventral temporal and posterior parietal cortices and the medial temporal lobe. One consequence was incomplete coverage of prefrontal cortex (PFC), meaning that stimulus representations in PFC could not be examined. Anatomical images were acquired with T1-weighted sequence (TR/TE = 11/3.3 ms, TI = 1,100 ms, 172 slices, flip angle = 18°, 1 mm3 resolution).

fMRI Data Analyses

fMRI Data Pre-processing

Pre-processing of anatomical and functional images was carried out using BrainVoyager (Brain Innovations) and custom Matlab scripts. Pre-processing included co-registration of functional scans to each individual's anatomical scan, slice-time correction, motion-correction, high-pass filtering (cut-off: 3 cycles/run), transformation to Talairach space (Talairach & Tournoux, 1988) and normalization (Z-scoring) of the functional timeseries data within each voxel for each run. Functional data from immediate repeat trials were removed from all multivariate pattern analyses

(MVPA). BOLD data for subsequent MVPA were extracted by snipping out the z-scored functional timeseries data for each stimulus trial from the 3rd and 4th TRs after stimulus onset (i.e., the period 4-8 seconds after stimulus onset) and averaging over the two datapoints.

Definition of ROIs

Retinotopic mapping was performed to define visual areas V1, V2v, V2d, V3v and V3d (Engel et al., 1994; Sereno et al., 1995). Data were collected in one or two scans per participant, using a flickering checkerboard wedge (8Hz flicker, 60^0 of polar angle) alternately presented at the horizontal and vertical meridians (20 s duration at each presentation). Because we did not collect functional localizer data for area LOC, we took an approximate — and therefore conservative — approach of defining a spherical ROI of radius 7mm centered upon the mean of the Talairach coordinates in each of left and right LOC reported by a set of 7 studies (Epstein, Higgins, Parker, Aguirre, & Cooperman, 2006; Grill-Spector, 2003; Grill-Spector et al., 1998; Large, Aldcroft, & Vilis, 2005; Lerner, Hendler, Ben-Bashat, Harel, & Malach, 2001; Song & Jiang, 2006; Xu, 2009). Coordinates were converted to from MNI to Talairach where necessary, yielding mean values in Talairch coordinates of Left LOC center [-44 -70 -4] and right LOC center [43 -67 -4]. Voxels were further screened for inclusion into each ROI (V1 through LOC) by taking the functional data from the ten experimental runs and performing a simple contrast of stimulus on versus stimulus off, testing against a liberal threshold of p=0.05, uncorrected. In all ROI-based multi-variate analyses, data from left and right hemispheres were combined into a single ROI, but the dorsal and ventral portions of areas V2 and V3 were kept separate.

Multivariate Pattern Analyses (MVPA)

After pre-processing, classification analyses were carried out using the libSVM software package (Chang & Lin, 2011) publicly available at https://www.csie.ntu.edu.tw/~cjlin/libsvm/. We used default parameters (e.g., cost parameter =1) and a linear kernel. For multi-class classification problems, libSVM uses a one-versus-one method, the performance of which is comparable to a one-versus-all method (Hsu & Lin, 2002). All classifier analyses were performed on each session in each subject individually; reported classifier accuracies are averaged over both sessions in each subject unless otherwise indicated. To ensure that over-fitting did not contribute to classifier performance we used hold-one-out cross-validation: classifiers were trained with BOLD data from nine (all but one) runs and tested with the tenth (held-out) run, with the process repeated ten times such that each run served as the test set once.

Comparing Feature- and Conjunction-Coding: The Feature Conjunction Index (FCI)

In order to determine how the relative levels of feature-based versus conjunction-based knowledge varied across brain regions, we devised a novel measure – the FCI – by placing classifier accuracies in a ratio (Figure 1). Positive FCI values indicate conjunction-coding, negative values indicate feature-coding, and zero values – provided classifier performance is above chance; which we ensured was the case by screening voxels or ROIs according to classifier performance – likely indicate a transition zone in which neither feature- nor conjunction-coding is strongly dominant. (See *Results* and Figure 2 for a demonstration and discussion of these properties of the FCI).

Examining the Coding of Mid-level Conjunctions

In addition to measuring decoding accuracy for simple features and for four-featured conjunctions, we also assessed 4-way classification of 'mid-level' conjunctions, i.e., combinations of just two of the four binary features possessed by each stimulus. There are six possible mid-level conjunctions: Feature 1 with Feature 2 ('Global Shape'), Feature 3 with Feature 4 ('Texture'), Feature 1 with Feature 3 ('Right Component'), Feature 2 with Feature 4 ('Left Component'); and two less plausible, unnameable conjunctions: Feature 1 with Feature 4, and Feature 2 with Feature 3. The mid-level conjunction classification accuracy allowed us to construct two further indices, by entering it into two ratios, (1) comparing feature versus mid-level conjunction knowledge in a *Feature vs. Mid-Level*

Conjunction Index (FMI), and (2) comparing mid-level conjunction versus whole-object conjunction knowledge in a *Mid-Level vs. Whole-Object Conjunction Index* (MCI). These indices were constructed analogously to the FCI (Figure 1). Calculation of FMI was performed separately for each unique, mid-level conjunction. For each mid-level conjunction, feature-classifier outputs were used to predict mid-level conjunction accuracy on a trial by trial basis, using only the two features that comprised the mid-level conjunction. Next, the predicted mid-level accuracy was compared with the empirical mid-level accuracy in a log ratio. The mean FMI value was an average over the four plausible mid-level conjunctions. For calculation of MCI values, we took the four plausible mid-level conjunction accuracies (i.e., outputs of four of the 4-way, mid-level conjunction classifiers) and combined them into two pairs to make two separate predictions for the whole-conjunction accuracy (Feat1-Feat2 combined with Feat3-Feat4; and Feat1-Feat3 with Feat2-Feat4), on a trial-by-trial basis. Reported MCI values were obtained by averaging across the two predictions and comparing this with the empirical whole-object conjunction accuracy in a log ratio.

Screening to Remove ROIs with Chance-level Classifier Accuracy

For both ROI-based and searchlight analyses, before computing the Feature Conjunction Index (FCI) we screened out ROIs in any session in any subject in which accuracy did not exceed chance for either feature-based or object-based classification. Analogously, for the FMI and MCI, we screened out any ROIs in which accuracy did not exceed chance for any of the classifiers contributing to the ratio. Chance level was determined using a binomial test, with binomial distribution parameters $n = 320$ trials, $p = 0.5$ for the two-way feature classifiers and $p = 0.0625$ for the sixteen-way object classifier, and with a statistical threshold $\alpha = 0.05$. The alpha level was adjusted for multiple comparisons by the Sidak method of assuming independent probabilities: $\alpha SID = 1 - (1 - \alpha)^{(1/n)}$ (where n=4 in the case of the 4 feature-classifiers). For the object-level object classifier, $\alpha SID = 0.05$ because there is only one classifier. For each of the 4 feature classifiers, $\alpha SID = 0.0128$ (i.e., $\alpha SID = 1 - (1 - \alpha)^{(1/n)}$, where $\alpha = 0.05$ and n=4). Because above-chance performance in either the object-classifier or any of the 4 feature classifiers qualified an ROI for inclusion, Sidak correction among the feature classifiers ensured that screening was unbiased with respect to feature-based versus object-based classifier accuracy. Binomial p values and Sidak adjustment were computed analogously for the FMI and MCI indices, according to the number of classifier outcomes and the number of comparisons in each case. The accuracy screening procedure ensured that ROIs with below-chance performance in all classifiers — from which computed FCI values are meaningless — were removed from the analysis. We did not correct for multiple comparisons when performing the binomial screening tests, because the outcomes of these tests were not our results of interest, they merely served to remove noise from the measure of interest, the FCI. In the searchlight FCI analyses, screening resulted in the removal of many centroid voxels. In the ROI-based analyses, for the FCI, screening resulted in the removal of one datapoint from a single session in one subject, in region LO (but removal of this datapoint did not significantly affect the results); for the FMI, screening resulted in the removal of several datapoints in region LO, described in the *Results*; and for the MCI, screening resulted in the removal of no datapoints.

ROI-based Statistical Tests

Differences in classifier accuracy per se were not of primary interest; our main goal was to calculate FCI by placing classifier accuracies in a ratio, in order to determine how the relative levels of feature-based versus conjunction-based knowledge varied across brain regions. Therefore, we examined differences in classifier accuracy only to provide preliminary descriptive characterization of the data, to verify that classifier performance was above chance, and to investigate whether there was an effect of scan session on the multivariate results. To test for differences in classifier accuracies across ROIs and sessions, accuracy scores for both feature- and object- classifiers in each session and subject separately were transformed into log likelihood ratios (log odds; $LLR(Acc) = \ln(Acc/(1-Acc))$. Log odds accuracy values for both feature- and object-classifiers were submitted to a two-way repeated measures ANOVA with Scan Session (1,2) and ROI (V1, V2v, V2d, V3v, V3d, LOC) as factors. Because this ANOVA revealed a significant effect of ROI for both feature- and object- classifier accuracy, we checked for adequate classifier performance in the lowest accuracy ROI (area LOC) by comparing LLR(acc) in LOC to chance performance (for features, chance LLR(acc) = 0; for objects, chance LLR(acc) = -2.71) via a one-sample t-test (one-tailed, alpha = 0.05). Ninety-five percent Confidence Intervals (CIs) for classifier accuracies and FCI (Tables 2-4), were determined by 10,000 iterations of bootstrap resampling with replacement. Resampling of classifier accuracy was conducted within-subjects, separately for each classifier and hold-out run, and these resampled values were averaged over the ten hold-out runs. To compute classifier accuracy CIs, the classifier accuracies were averaged across the two sessions and all subjects for each iteration and compiled into a distribution of mean classifier accuracies. To compute CIs for the FCI mean, an FCI value was computed from the classifier accuracies for each iteration (separately for each classifier and ROI) and averaged across the two sessions and across subjects, before being compiled into a distribution of mean FCI values. Finally, 95% CIs were drawn from these distributions by taking the 250th and 9750th ranked values. Comparison of CIs between pairs of ROIs provides an assessment of differences in classifier accuracy or FCIs between ROIs.

The FCI is a log ratio centered on zero, therefore we performed no further transformation before submitting it to a two-way repeated measures ANOVA with factors Scan Session (1,2) and ROI (V1, V2v, V2d, V3v, V3d, LOC). Because there was a significant effect of ROI, we performed Sidak-adjusted pairwise comparisons to test for differences between ROIs. For the FCI, ninety-five percent CIs (Figure 2; Table 4) were determined by 10,000 iterations of bootstrap resampling with replacement, providing a secondary assessment of differences between ROIs. We also performed two-way repeated measures ANOVAs on the FMI and MCI values, but in the case of the FMI we tested only 5 ROIs, excluding LOC, because of missing datapoints for LOC following screening.

Searchlight Analyses

To assess the relative dominance of feature- versus conjunction-coding throughout all of visual cortex we computed FCI using a searchlight approach (Kriegeskorte, Goebel, & Bandettini, 2006). The imaged volume in each session in each subject was first screened using a subject-specific gray-matter mask encompassing occipital, temporal and posterior parietal cortices; the approximate volume encompassed for each participant can be seen in Figure 4. A sphere of radius 5 functional voxels was sampled around each voxel in this volume (any sphere containing fewer than 100 voxels falling within the gray-matter mask was excluded); feature- and object-classifiers were trained on the BOLD data from the spherical ROI, and the resulting FCI was recorded at the centroid voxel in the map. To create the group average FCI map in Figure 4, we first took the FCI map for each session and subject performed spatial smoothing using a Gaussian kernel with FWHM of 2. The spatially smoothed FCI value at each voxel was computed by performing spatial smoothing of (1) object classifier accuracy and (2) feature-predicted object classification accuracy, and placing the two smoothed values into a log ratio, as in Equation 1. At each voxel, any neighboring voxels that were disqualified from inclusion in the analysis by our standard criteria — falling outside of the anatomical mask, possessing fewer than 100 voxels in their associated spherical ROI, or yielding above chance classifier accuracy for neither features nor objects — were treated as missing values in the Gaussian averaging calculation. Next, for each subject, the smoothed maps were averaged over two sessions. (The collapsing of data across two sessions was justified because in the ROI-based analyses we found no effect of session on FCI, nor any interaction of session with ROI; see *Results*. In addition, for the single-session FCI maps generated by searchlight analysis, we conducted a group-level t-test at each voxel for a difference of FCI across sessions and found no voxels passing an FDR-corrected threshold of alpha=0.05). Finally, the group average FCI map (Figure 4) was constructed by averaging over all subjects' smoothed session-averaged FCI values at each voxel, with the constraint that for a voxel to appear in the map it had to possess a numeric FCI value (rather than a missing value, indicating that the voxel had been screened out) for at least 5 of the 8 subjects.

Group-level Statistical Tests on Searchlight Data

To find cortical locations of reliably non-zero FCI values — either positive indicating conjunction-coding or negative indicating feature-coding — we performed a group-level t-test at each voxel in the searchlight analysis, comparing the group mean FCI to zero. To do so, we took FCI maps from individual subjects and sessions that had been spatially smoothed using a Gaussian kernel with FWHM of 2 functional voxels, and averaged over the two sessions for each subject, as described above. We tested only voxels that were associated with numeric FCI values for all 8 subjects (i.e., voxels that did not contain an empty value after spatial smoothing, owing to disqualification during the screening process). We used a False Discovery Rate (FDR)-corrected alpha level of 0.05 (two-tailed) against a t-distribution with 7 degrees of freedom. Anatomical labels for identified sites in cortex were derived from the Automated Talairach Atlas available at www.talairach.org (Lancaster et al., 2000).

Quantifying the transition from feature- to conjunction-coding

In order to quantify the transition from feature to conjunction-coding along the ventral and dorsal pathways, we examined the FCI as a function of location along each pathway. To specify the location of a voxel in the ventral and dorsal pathways, we first defined three vectors in Talaraich co-ordinates: (1) a 'Posterior Ventral' vector with its origin in the occipital pole (Tal co-ords, L: [-8 -101 -6]; R: [8 -101 -6]) extending to the center of LOC (Tal co-ords, L: [-44 -70 -4]; R: [43 -67 -4]); (2) an 'Anterior Ventral' vector with its origin at the center of LOC, extending to the anterior tip of the temporal pole (Tal co-ords, L: [-35 28 -29]; R: [35 28 -29]); and (3) a 'Dorsal' vector with its origin in inferior posterior occipital cortex (Tal Co-ords L: [-10 -99 -14]; R: [10 -99 -14]) extending to the most superior/anterior point of the dorsal pathway contained in the scanned volume, in

Brodmann Area 7 (Tal Co-ords L: [-14 -67 53]; R: [14 -67 53]). The goal was to project the Talairach co-ordinates of each voxel in the dorsal and ventral pathways onto the three vectors we defined, to produce a single, scalar metric specifying the location of each voxel along each pathway. In order to include only voxels in the appropriate cortical regions for each vector (e.g., to exclude anterior voxels from the 'Posterior Ventral' vector, and ventral voxels from the 'Dorsal' vector), we defined a bounding box around each vector, outside of which voxels were excluded from the analysis (see Figure 6). In all cases, the bounding box for each hemisphere terminated at X=0 (the midline). For the Posterior Ventral vector, the boundaries of the box were: X=0 to X = lateral extent of Talairach bounding box; Y = -55 to the posterior extent of Talairach bounding box; Z = 7 to Z = inferior extent of Talairach bounding box. For the Anterior Ventral vector, the bounding box dimensions were: X=0 to X = lateral extent of Talairach bounding box; Y = -70 to the anterior extent of Talairach bounding box; Z = 7 to Z = inferior extent of Talairach bounding box. For the Dorsal vector, the bounding box dimensions were: X = 0 to X = lateral extent of Talairach bounding box; Y = -57 to the posterior extent of the Talairach bounding box; Z = -14 to the posterior extent of the Talairach bounding box. In addition, for the Posterior Ventral Vector, we excluded any voxels with a scalar projection value beyond the end of the vector, which terminated at the center of LOC (i.e., any voxels within the bounding box located at a more extreme point along the vector than the center of LOC).

Having derived a metric specifying the location of each voxel in each of the three pathways, we plotted FCI values as a function of voxel location. FCI values were not spatially smoothed, but were averaged over the two sessions for each voxel, within each subject. Separate plots were made for each subject and hemisphere, for each of the three defined pathways (posterior Ventral, Anterior Ventral and Dorsal). For each plot, we assessed the correlation between voxel location and FCI, and found the best fitting straight line describing the relationship by a least-squares method. The best fitting straight lines for all subjects and hemispheres in all three pathways are shown in Figure 6. We assessed differences in the slope of the best-fitting straight lines for the three pathways by performing a 2x3 repeated measures ANOVA with factors Hemisphere (Left, Right) and Pathway (Posterior Ventral, Anterior Ventral, Dorsal).

Evaluating the Properties of the Feature Conjunction Index (FCI) with Synthetic Data

Construction of Synthetic Data

To evaluate the FCI metric we created synthetic BOLD datasets. We used two different templates – feature-coded and conjunction-coded – to generate the signal that defined distinct patterns for the 16 stimuli in the set. In the feature-coded template, artificial voxels were switched 'on' or 'off' according to the presence or absence of a particular value of one of the four binary features, with all feature values represented by an equal number of voxels. In the conjunction-coded template, artificial voxels were switched 'on' or 'off' according to the presence or absence of a specific object stimulus from among the set of 16, with all objects represented by an equal number of voxels. All datasets possessed 256 voxels in total. We generated synthesized datasets in two ways: (1) by selecting a signal value from a range of magnitudes (between 0.01 and 0.5), applying this signal to 'active' voxels according to the template, and superimposing this pattern of activation on top of a constant background of uniform random noise (range 0 to 1, standard deviation 0.289), resulting in two families of datasets possessing signal-to-noise ratios (SNR) ranging from 0.0012 to 3, and (2) by using a constant signal value of 1, applying this signal to 'active' voxels according to the template, and superimposing this pattern on top of a background of uniform random noise whose amplitude was systematically manipulated (the range of the noise varied from a minimum range of 0-2, with standard deviation of 0.5774, to a maximum range of 0-30, with standard deviation 8.66), resulting in two families of datasets (feature- and conjunction-coded) that possessed signal to noise ratios (SNR) ranging from 0.0133 to 3. For both data synthesizing methods, data were normalized on a ‘per run’ basis exactly as empirical BOLD data were normalized before analysis (grouping 'trials' into sets of 32 containing two presentations per stimulus – a single scanner run – and normalizing across all trials in a run for each voxel).

Assessment of FCI

Our goals were (1) to assess the intuition that feature-coded activation patterns produce negative FCIs and conjunction-coded patterns produce positive FCIs, when classifier performance is above chance, and (2) to examine the extent to which qualitative or quantitative shifts in FCI are produced by varying levels of noise. Following identical procedures to those used for empirical BOLD data, taking 100 synthetic datasets from each template – feature-coded and conjunction-coded – we ran feature- and conjunction-level classifiers and computed FCI values from the classifier outputs. We then plotted FCI values and classifier accuracies against SNR, for both data templates. Critically, we aimed to evaluate the FCI metric against the results from datasets that produced classifier accuracies in line with the range of accuracies seen in our empirical data. This is important because FCI values necessarily tend to zero when both feature- and object-level classifiers either know nothing (below chance performance) or have perfect knowledge (approaching 100% accuracy). We emphasize that, in our empirical data, classifier accuracies never approached ceiling, and below-chance classification accuracies always disqualified an ROI from inclusion in the assessment of FCI.

Results

FCI Reflects the Relative Contribution of Feature- versus Conjunction-Coding in a Cortical Region

As seen in Figure 2, analyses of synthetic data constructed according to a feature-code and a conjunction-code yielded FCI values that confirmed the interpretation of FCI outlined in the *Methods*. Focusing on the datapoints within the gray boxes (for which mean classifier accuracies fell within the range observed in ROI-based analyses of the empirical data), when FCI values from each set of 100 simulations were compared to zero with a one-sample t-test ($\alpha = 0.05$), it was confirmed that all Feature-Coded datasets produced negative FCI values and all Conjunction-Coded datasets produced significantly positive FCI values. We emphasize two key features of these simulation results. First, positive FCI values were never produced by a feature-coded template, and negative FCIs were never produced by a conjunction-coded template. Second, for both templates, increasing noise tended to push the FCI toward zero, making it less negative for feature-coded data, and less positive for conjunction-coded data. However, in these simulations, FCI values derived from above-chance, empirically-plausible classifier accuracies were always statistically distinguishable from zero (i.e., significantly negative or positive). Taken together, the simulations demonstrate that a region containing only object-sensitive voxels cannot produce a negative FCI and a region containing only feature-sensitive voxels cannot produce a positive FCI, precluding qualitatively misleading results. When the FCI takes near-zero values, in the presence of above-chance classifier accuracy, this likely represents a transition zone containing some mixture of feature and conjunction sensitive voxels, such that neither coding type is dominant. In the presence of high noise levels, zero FCIs are harder to interpret, because they may reflect noise rather than a balance of feature and conjunction coding. It is for this reason that, in the empirical analyses (both ROI-based and spherelight), we screened out any ROIs for which classifier accuracy did not exceed chance. Despite this caveat regarding near-zero FCIs, it is nonetheless instructive that, even at high noise levels, numerically positive FCIs never emerged from a feature code and numerically negative FCIs were never produced by a conjunction code. The simulations thus reinforce the a priori intuition that the FCI indicates whether the voxel-level population code is relatively dominated by a feature-code (negative FCI) or a conjunction-code (positive FCI).

An intuitive understanding of the pattern of FCI values derived from synthetic data can be gained by considering how individual voxel activations map onto category identities, for each underlying coding type and classifier type. All classifiers used a linear kernel. First, consider that for hypothetical activation patterns with zero noise, both feature- and conjunction-knowledge would be perfect regardless of the underlying code (feature-based or conjunction-based), and FCI values would be zero. However, empirical BOLD data contain noise, which produces incomplete knowledge. In the case of noisy feature-coded data, reliable information is present for some but not all of the features (i.e., noise obscures the signal in some but not all of the feature-sensitive voxels). Each feature classifier must map feature-sensitive voxels onto feature categories, but these mappings are independent for the different categories: different voxels carry the information for each category. Although some voxels' signal is obscured, reducing the feature classification accuracy for the feature categories coded by those voxels, the feature categories for which reliable information is present can still be classified because mappings from the reliable voxels to feature categories are unaffected by noise in the unreliable voxels. In contrast, the conjunction classifier must map feature-sensitive voxels onto object categories, which requires combining information from all features, therefore all voxel-to-category mappings are affected by the loss of information in some voxels. Consequently, object classification accuracy drops further than feature classification accuracy, and FCI is negative. In the case of noisy conjunction-coded data, the reverse scenario applies. Since the conjunction classifier learns independent mappings of conjunction-sensitive voxels to object categories (separate voxels in each case), noise on some but not all voxels reduces accuracy for some but not all categories. In contrast, the feature classifier must perform combinatorial mappings of conjunction-sensitive voxels onto feature categories, so the presence of noise on some voxels affects all voxel-to-feature category mappings. Hence, noise affects feature classifier accuracy more than conjunction classifier accuracy, producing positive FCI values.

What does this imply for the nature of the neural code? Positive FCIs are taken as indicating conjunction-coded activation patterns. In our simulations, this amounted to voxels in which the signal varied with the presence or absence of an object, without regard to its component features. That is, the intersection of the features comprising the object created a wholly new pattern, distinct from the pattern due to other objects sharing some of those features. A conjunction-code is thus a non-linear combination of features for which the whole is greater than the sum of the parts.

A Transition from Feature- to Conjunction-Coding in Occipito-Temporal Cortex

We investigated whether early visual cortex employs feature-coding or conjunction-coding, and whether one scheme transitions to the other with progression toward the temporal lobe, by examining FCI in a series of functionally or anatomically defined ROIs: V1, V2, V3 and lateral occipital cortex (LOC). Taking stimulus-evoked activation patterns from across each ROI, we trained the classifiers using hold-one-out cross-validation and computed FCI for each subject and session separately (Figure 3; Table 4). FCI differed significantly across ROIs ($F(5,35) = 15.78$, $p < 0.001$, $\eta^2=0.693$), but not across sessions ($F(1,7) = 0.004$, $p = 0.95$, $\eta^2=0.001$), with no interaction between Session and ROI ($F(5,35)= 0.133$, $p = 0.984$, $\eta^2=0.019$). Ordering the ROIs as [V1, V2v, V2d, V3v, V3d, LOC] revealed a significant linear trend ($F(1,7) = 37.34$, $p < 0.001$; $\eta^2= 0.842$), suggestive of a transition from feature-coding toward conjunction-coding along early stages of the visual pathway. Regions V1 through V3 exhibited negative FCIs indicative of feature-code dominance, while LOC revealed a numerically positive FCI that was significantly greater than in all other ROIs (revealed by non-overlapping 95% CIs, Table 4). In our analysis of synthetic data, neither purely feature-based nor purely conjunction-based data (at SNR levels that yielded classifier accuracy within the range observed in the empirical ROI-based analyses) produced zero FCIs. We therefore interpret the near-zero, numerically positive FCI for LOC in terms of a hybrid code that contains both feature- and conjunction-coding, in contrast to the strong feature-coding detected in early visual cortex. In favor of this interpretation, we note two points. First, both feature and object classifier accuracy exceeded chance in LOC (see Tables 2 and 3). Given this above-chance classifier performance, and the simulation results reported above, it would not be possible to derive a numerically positive FCI if the true underlying code in LOC were as strongly feature-based as in regions V1 through V3 (indicated by their very negative FCI values). Second, region LOC showed a significantly positive

FMI for the mid-level conjunction of global shape (reported below), indicating that the quality of BOLD in LOC, in this dataset, is sufficient to produce non-zero indices when the information present in the activation patterns warrants such values.

Feature- and Conjunction-Coding Throughout Visual Cortex

In order to assess feature- and conjunction-coding in cortical representations of objects beyond region LOC, and in the dorsal visual pathway, we examined all of visual cortex using a searchlight approach (Kriegeskorte et al., 2006). At each spherical ROI, we performed classifier analyses (using hold-one-out cross-validation and screening out spheres in which classifier accuracy did not exceed chance) and computed the FCI, mapping the FCI value back to the centroid voxel of the sphere. In the group-averaged FCI map (Figure 4), occipital regions exhibit the most negative FCIs (green) indicative of feature-coding. With progression into regions anterior and superior to the occipital pole, the FCI first becomes less negative (blue) – suggesting a transition, or ‘hybrid’ region – and then becomes positive (orange/yellow) in occipito-temporal and posterior parietal regions, indicating the emergence of conjunction-coding in both ventral and dorsal pathways. Examination of FCI maps in individual subjects revealed the same pattern in every subject in both sessions: strongly negative FCIs in occipital regions with a transition to positive FCIs toward temporal and parietal regions (Figure 5).

Quantifying the Transition from Feature- to Conjunction-Coding

Next, we sought to quantify the relationship between cortical location and FCI in both ventral and dorsal pathways. To do so, we devised a metric to specify the location of the voxels in each pathway by defining three vectors in Talaraich co-ordinates: a '*Posterior Ventral*' vector with its origin in the occipital pole extending to the center of LOC; an '*Anterior Ventral*' vector with its origin in LOC extending to the anterior tip of the temporal pole; and a '*Dorsal*' vector with its origin in inferior posterior occipital cortex extending to the most superior/anterior point of the dorsal pathway in the scanned volume (in Brodmann Area 7). For each vector we defined a bounding box around the vector to constrain the anatomical region from which voxels were drawn (Figure 6) and projected the Talairach co-ordinates of voxels within the box onto the vector, yielding a scalar value for each voxel that specified its location along the vector. Finally, for each subject, for all three vectors in each hemisphere separately, we computed (i) the correlation between the location of a voxel and the FCI of the spherical ROI centered on that voxel, and (ii) the slope of the best fitting regression line relating voxel location to FCI (Figure 6). The correlation between location and FCI was positive and highly significant in all 8 subjects in both hemispheres for the *Dorsal* vector ($p< 0.0001$) and the *Posterior Ventral* vector ($p<0.01$), reflecting a robust transition from feature-coding at the occipital pole to representations more dominated by conjunction-coding in lateral occipital and superior parietal cortices, respectively. For the *Anterior Ventral* vector, the correlation was positive and significant ($p<0.01$) in both hemispheres for 5 out of 8 subjects; in 1 subject (yellow in Figure 6) the left hemisphere was negatively correlated (a decrease in FCI with anterior progression; $p<0.05$) and the right was positively correlated ($p<0.001$); in the 2 remaining subjects (black and magenta in Figure 6), the left hemisphere was significantly negatively correlated ($p<0.0001$) and the right was not correlated ($p>0.2$). The slopes of the best fitting regression lines differed for the three vectors ($F(2,14)=23.71$, $p<.001$), but did not differ by hemisphere ($F(1,7)=.026$, $p=.877$), with no hemisphere by vector interaction ($F(2,14)=2.889$, $p=.089$). Slopes for the *Anterior Ventral* vector were smaller than for the *Posterior Ventral* ($p < 0.001$) and *Dorsal* ($p < 0.0001$) vectors, which did not differ from each other ($p = 0.62$). These results suggest that, for the present stimulus set, the greatest transition from feature- to conjunction-coding occurs in posterior regions, in both ventral and dorsal pathways. A possible reason for the shallower transition toward conjunction-coding in the anterior ventral pathway is that the object-level conjunctions comprising these simple, novel objects may be fully specified in relatively posterior sites, just beyond the occipito-temporal junction (Figures 4 and 7). Indeed, the stimuli with which Erez et al. (2015) revealed conjunction coding in anterior temporal regions were 3-dimensional, colored objects that likely better engaged anterior visual regions.

Cortical Sites of Extreme Feature- and Conjunction-Coding

Finally, to search for cortical sites demonstrating statistically reliable extremes of feature- or conjunction-coding, we compared the group mean FCI at each voxel to zero (two-tailed t-test, False Discovery Rate (FDR) corrected; positive t-values indicate conjunction-coding and negative t-values feature-coding). This analysis assumes anatomical and functional correspondence of points in Talairach space across subjects; it is therefore conservative, particularly for conjunction-coding, the cortical sites of which are likely more widely and variably distributed across subjects. Nonetheless, we revealed a large occipital region of feature-coding along with multiple conjunction-coding sites throughout occipito-temporal, ventral temporal and parietal cortices, extending into the parahippocampal gyrus, medial temporal lobe and anterior temporal pole (Figure 7). We note the lower prevalence of above-threshold voxels contributing to each subject's FCI map in regions toward the anterior temporal lobes (see Figure 5), and we therefore interpret our findings in these regions with caution; in contrast, data in the occipital, posterior temporal and posterior parietal lobes were much less sparse and so findings in these regions can be interpreted with greater confidence. Most important is the overall pattern at the group-level: in line with a *feature-to-conjunction transition hypothesis*, all conjunction-coding sites were located anterior or superior to the feature-coding sites, which were confined to the occipital lobe (excepting a single more anterior feature-coding voxel at [10 -51 5], in the inferior posterior cingulate).

No Effect of Visual Search Training on Feature- or Conjunction-Coding as Measured by FCI

We detected no effect of visual search training on the prevalence of feature versus conjunction coding, as measured by the FCI, in participants' cortical activation patterns. As reported above, there was no main effect of Session and no Session*ROI interaction on the critical FCI measure (where Session is the factor partitioning data collected before versus after visual search training). Because the visual search task required object individuation, any observed change in neural representations was expected to increase conjunction coding. There are several possible explanations for the absence of an observed training effect on the neural representations. First, in our training procedure, daily sessions were terminated for each subject when the RT-set size slope appeared to be approaching an asymptotically low value (mean 11.1 days; range 7 - 15 days). This relatively short training duration may have been insufficient to produce full unitization of the conjunctive stimuli, a notion in line with the fact that RT-set size slopes were still significantly positive in the final training session (Table 5). Second, the stimuli were perceptually simple and devoid of semantics, and the task did not involve naming; thus, our training may have failed to engage the learning mechanisms known to influence neural representations in visual cortex for richer, more meaningful objects (e.g., Folstein, Palmeri, & Gauthier, 2013; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). A third possibility is that cortical representations did change with training, but the data were too noisy to allow detection of these changes at the individual subject level, and the sites at which changes occurred varied in location across participants, obscuring findings at the group level.

Coding of Mid-Level Conjunctions

We examined the extent to which ROIs in the ventral and dorsal pathways coded for simple features versus mid-level conjunctions (FMI, Figure 8), and coded for mid-level conjunctions versus whole-object conjunctions (MCI, Figure 9). We found the same trend as revealed by the FCI analysis: both FMI and MCI measures took negative values in posterior regions such as V1 (indicating that feature-coding dominates over mid-level conjunction-coding, and mid-level conjunction-coding dominates over whole-object conjunction-coding), but increased toward positive values in more anterior regions such as V3d and LOC.

Statistical assessment of FMI by ANOVA excluded LOC because of missing cells. Specifically, screening (to remove any session-subject-ROI datapoint for which no classifier exceeded chance as determined by a binomial – see Methods) resulted in the removal of several datapoints in LO only, such that Mid-Level Conjunctions 1 and 3 were determined on the basis of 6 and 7 subjects, respectively, and the mean over all conjunctions was determined on the basis of 6 subjects (i.e., only those individuals who contributed a datapoint for all 4 mid-level conjunctions). ANOVA performed on the remaining ROIs revealed that FMI differed significantly across ROIs ($F(4,28) = 21.07$, $p < 0.001$, $\eta^2=0.751$), and across sessions ($F(1,7) = 11.96$, $p = 0.011$, $\eta^2=0.631$), with no interaction between Session and ROI ($F(4,28)= 1.54$, $p = 0.219$, $\eta^2=0.180$). The small but significant difference across sessions was due to a decrease of 0.015 in FMI from the first to second scan sessions, i.e., a change after visual search training toward feature-coding and away from mid-level conjunction coding. Since visual search training focused on the unique identity of the whole object, it is not clear what should have been expected for any shift in neural coding of mid-level conjunctions. This shift toward more negative FMIs might result from increased attention to individual features over mid-level conjunctions after training, but the shift was small in magnitude.

Statistical assessment of MCI by ANOVA revealed that MCI differed significantly across ROIs ($F(5,35) = 17.79$, $p < 0.001$, $\eta^2=0.718$), but not across sessions ($F(1,7) = 0.026$, $p = 0.876$, $\eta^2=0.004$), with no interaction between Session and ROI ($F(5,35)= 0.72$, $p = 0.613$, $\eta^2=0.093$).

Together, the FMI and MCI results yield three notable insights. First, the FMI analysis demonstrates that when simple conjunctions (rather than whole-object conjunctions) are pitted against the basic features, the zero point of the transition from feature-coding to conjunction-coding occurs earlier in both visual pathways, near area V2. Second, the combination of FMI and MCI results for area V3d suggest that it codes primarily for mid-level conjunctions in the present stimulus set: in V3d, the FMI (features versus mid-level conjunctions) was significantly positive (one-sample t-test compared to zero, $p<0.001$) whereas the MCI (mid-level versus whole-object conjunctions) was significantly negative (one-sample t-test, p = 0.0016), suggesting that mid-level conjunction-coding dominates over both feature-coding and whole-object conjunction-coding. Third, the FMI values for area LOC suggest that the extent to which this region codes for conjunctions relies heavily upon its coding of shape. As seen in Table 7, which shows FMI values for individual mid-level conjunctions, LOC yielded a significantly positive value only in the case of Conjunction 1 – the combination of Left and Right outlines corresponding to the global form of the stimulus (t-test, p=0.0019). (We note that, even though the contour features each possess their own 'global' shape, a positive FMI value for Conjunction 1 nonetheless indicates that the brain region codes for the intersection of the two contour features. If the activation patterns comprised only a linear combination of the separate representations of each contour feature, the conjunction classifier would show no performance advantage over the feature classifiers.) In contrast, for Conjunction 2 (the combination of two spatial frequencies, or 'texture'), LOC yielded a significantly negative FMI (t-test, p =0.0388) indicative of dominant feature-coding.

Discussion

Several recent fMRI studies have examined conjunction-coding in humans (Baeck, Wagemans, & Op de Beeck, 2013; Baumgartner et al., 2013; Erez et al., 2015; Macevoy & Epstein, 2009; Pollmann, Zinke, Baumgartner, Geringswald, & Hanke, 2014; Seymour et al., 2009, 2010; van den Honert, McCarthy, & Johnson, 2017; Zhang, Liu, & Xu, 2015). Two of these studies revealed conjunction-coding for combinations of static, abstract features (e.g., color, form) into visual objects, in the ventral pathway (Erez et al., 2015; Seymour et al., 2010). However, neither study was able to definitively rule out an explanation of the observed conjunction-code in terms of feature-coding combined with saturation of the BOLD signal; that is, because both studies presented the critical conjunctions by temporally interleaving subsets of features that comprise the conjunction, rather than by presenting the whole conjunction simultaneously, the presence of neurovascular nonlinearities (i.e. saturation of BOLD) for some features more than others may have given the appearance of a conjunction-code even in feature-coding voxels. Two further studies in this set (Baumgartner et al., 2013; Pollmann et al., 2014) used a method more similar to ours, measuring whether classification accuracy for two feature dimensions together was super-additive compared to the classification of each dimension alone. However – perhaps related to the use of multiple-object stimulus displays and a visual search task – these studies revealed conjunction-coding only in parietal regions. Moreover, no study of object processing that we know of has explicitly investigated the complement to conjunction-coding, namely, feature-coding – that is, the extent to which neural representations are *more informative* about individual features than conjunctions of those features (but see van den Honert et al., 2017, for a slightly different definition of feature-based representations in a study of scene-level conjunctions). Without directly comparable measures of feature-coding and conjunction-coding, the evidence for a transition from one to the other cannot be assessed. Therefore, three important questions remain unresolved: Is there evidence for a conjunction-based object code in the human ventral visual stream, when an explanation in terms of BOLD signal saturation is ruled out? If yes, is conjunction-coding dominant even early in the ventral stream or does it emerge along the pathway? If it emerges, at what point does the transition from feature- to conjunction-coding occur?

To answer these questions, we measured the relative dominance of feature- versus conjunction-coding throughout visual cortex by directly comparing evidence for the two coding schemes in each cortical region. Critically, evidence for both schemes was derived from a common neuroimaging dataset, acquired while participants viewed systematically constructed conjunctive visual stimuli. We revealed a transition from feature-dominated to conjunction-dominated coding, with progression from primary visual cortex into temporal cortex (Figures 3, 4, and 6). This provides the first direct evidence in humans for an object representation scheme in the ventral pathway that is characterized by a *transition from feature- to conjunction-coding*. Strikingly, the same shift from feature- to conjunction-coding was evident in the dorsal pathway, where it was as steep and robust as in the posterior ventral pathway (Figures 3 and 6).

The results are incompatible with a *global feature-coding hypothesis*, which predicts the feature-coding we observed in early regions but not the conjunction-coding that we saw in both parietal and anterior ventral sites (Figure 7). Similarly, the results are incompatible with a *global conjunction-coding hypothesis*, which predicts the conjunction-coding seen in downstream sites but cannot accommodate the findings in V1, where – regardless of whether feature-coding was compared to the coding of mid-level conjunctions (FMI) or whole conjunctions (FCI) – the index was always negative, reflecting more information about features than about the conjunctions of those features (Figure 8 and 9). Thus, the explicit measurement of feature-coding via negative FCI values was critical to distinguishing between the alternative theoretical accounts.

We revealed significant conjunction-coding in a range of sites bilaterally, including posterior parietal cortex, fusiform and parahippocampal gyri, medial temporal lobes and anterior temporal poles (Figure 7). These findings concur with prior reports of conjunction-coding throughout ventral visual stream and perirhinal cortex (Erez et al., 2015) and in parietal lobe (Baumgartner et al., 2013; Pollmann et al., 2014). In addition, we provided a systematic investigation of feature-coding across human visual cortex, revealing significant selectivity for features over conjunctions in retinotopically-defined V1 (Figures 8 and 9) and across posterior occipital regions more generally (Figure 7). For the present stimulus set, the transition from feature to conjunction-coding — the zero point of the FCI — occurred in the ventral pathway near the occipito-temporal junction, and in the dorsal pathway at approximately the superior border between Brodmann areas 18 and 19 of the occipital lobe (Figure 4).

We also examined coding for intermediate-complexity stimulus components, by pitting features against mid-level conjunctions (in the FMI) and mid-level conjunctions against whole-object conjunctions (in the MCI). These measures suggested that the transition from feature- to conjunction-coding occurs earlier in the visual pathway for simpler conjunctions, that V3d may preferentially code for mid-level conjunctions, and that conjunction-coding in LOC may be driven by coding of shape, or global form, in line with prior reports (Drucker & Aguirre, 2009; Malach et al., 1995; Ostwald et al., 2008). The mid-level conjunction findings also corroborated our interpretation of the main FCI measure, as discussed below.

The transition from feature- to conjunction-coding: gradual or abrupt?

Because feature-coding dominates in V1 and conjunction-coding sites exist in fusiform gyrus and posterior parietal lobe (Figure 7), we can conclude that somewhere between early visual areas and later stages of both the ventral and dorsal pathways a transition from feature- to conjunction-coding occurs. Figures 4, 5 (whole-brain analyses) and 6 (correlation analysis) imply a gradual transition. However, it is not possible to conclude definitively from the FCI that the transition from feature- to conjunction-coding in the brain is gradual, rather than abrupt. A graded change in the FCI is somewhat guaranteed for the whole-brain, spherelight analyses by the spatially overlapping spherical ROIs, and for the correlations by the assumed linear relationship between FCI and voxel location. We note that this guarantee does not apply to the ROI results in Figure 3, which reflect the same gradual transition, because we did not perform any spatial averaging of the data in this analysis. However, even in this case, it is possible that a reduction in SNR from V1 toward higher visual regions (V3, LOC) induced FCI values to increase toward zero, without an underlying gradual shift from feature-coded toward conjunction-coded representations (see simulations, Figure 2). Thus, it may be that the cortical code shifts abruptly from feature- to conjunction-based somewhere between V3 and IT, and the appearance of a gradual transition from V1 through LOC is caused by increasing noise. However, we think this unlikely given the FMI values that were derived from the same BOLD data (Figure 8). Specifically, the FMI increased from negative values in V1, through zero in V2, to positive values in V3d. That is, for simpler conjunctions, a similarly gradual transition in coding occurs across the same brain regions, but the steady increase in FMI values does not stop at zero (as it would if driven by noise alone) but instead pushes through to positive, conjunction-coding values in V3d. Thus, for FMI, an explanation involving increasing noise with progression up the pathway cannot explain the pattern of results.

Conjunction-coding in the dorsal pathway

The emergence of conjunction-coding in the dorsal pathway was perhaps more unexpected than in the ventral pathway. However, this finding aligns well both with the documented role of parietal cortex in feature binding (Ashbridge, Walsh, & Cowey, 1997; Cohen & Rafal, 1991; Friedman-Hill, Robertson, & Treisman, 1995; Humphreys, Cinel, Wolfe, Olson, & Klempen, 2000; Muggleton, Cowey, & Walsh, 2008) and with recent claims that the role of dorsal stream in vision extends beyond spatial processing or attentional binding. That is, our observation of object-specific coding in parietal cortex – elicited while participants discriminated highly similar visual objects – suggests that the dorsal pathway constructs content-rich, hierarchical representations containing information that is critical for object identification, in parallel with the ventral stream (Jeong & Xu, 2016; Konen & Kastner, 2008). Given that feedback from parietal cortex has been shown to shape visual representations in the ventral processing pathway (e.g., Hopfinger, Buonocore, & Mangun, 2000; Parks et al., 2015; Rowe, Stephan, Friston, Frackowiak, & Passingham, 2005), it is possible that interactions between dorsal and ventral regions increased the similarity of representations across the two streams.

V2: Coding of features and simple conjunctions

The negative FCI recorded for V2 suggests feature-coding (Figure 3), which might appear to contradict previous reports of conjunction-coding in this region for stimuli comprising combinations of simple features (Anzai et al., 2007; Seymour et al., 2010). However, negative FCI values in V2 do not rule out the existence of conjunction-based information altogether, they merely imply that any conjunction-code is relatively swamped by a stronger feature-code. Moreover, the complementary FMI metric that assesses coding of simpler conjunctions – similar in complexity to the stimuli of Anzai et al. and Seymour et al. – revealed near-zero (numerically positive) scores in V2. These FMI values imply an intermediate code in V2 that includes a substantial subset of voxels sensitive to mid-level components. A strength of the present method, whether used to compare features with mid-level conjunctions (FMI) or with whole-object conjunctions (FCI), is that it reveals the relative dominance of feature-based versus conjunction-based information in each case, rather than simply detecting the presence or absence of only one type of information.

Interpreting the FCI

An analysis of synthetic data supported our interpretations of the FCI metric (Figure 2). Across a range of SNR values for which classifier performance was within the range observed in empirical ROIs (and thus, also, above chance), negative FCIs were produced only by synthetic feature-coded data and positive FCIs emerged only from synthetic conjunction-coded data. These properties held true whether we synthesized data by holding noise constant and manipulating signal strength, or by holding signal constant and manipulating noise. Thus, the sign of the FCI reliably indicates the relative dominance of feature- versus conjunction-coding, and varying noise levels do not produce distortions in the FCI that lead to qualitatively erroneous conclusions. Accordingly, the method mitigates somewhat against the problem of varying noise levels that has complicated prior attempts to use fMRI to compare the neural code across diverse brain regions. For example, if a standard MVPA method detects greater classification accuracy for simple visual features in early visual cortex than in later visual regions, this could be because the neural representations in early visual regions exhibit stronger feature-coding, or because early visual cortex, situated peripherally, produces a BOLD signal with greater SNR. In the present method, because the feature- and conjunction-coding measures are placed in a ratio, and both measures are affected by the noise in each cortical region, the relative dominance of feature- versus conjunction-coding in a region maps consistently to negative versus positive FCI values in the face of varying noise.

In sum, the simulations demonstrate that the FCI can be used to determine which regions exhibit feature- versus conjunction-coding and approximately localize the region in cortex where the code transitions from one to the other, for a given stimulus set.

Potential influence of sensitivity to mid-level conjunctions on the FCI

Because the binary features used in our dataset included two shape outlines and two spatial frequencies, the intersection of any two of these features created incidental, novel shapes and textures that correspond to ‘mid-level conjunctions’ like those we decoded for entering into the FMI and MCI (for example, in Figure 1D: the global outline shared by stimuli 1, 3, 9 and 11, or the central texture shared by stimuli 1, 2, 5 and 6). Consider a brain region containing voxels sensitive to these mid-level conjunctions, rather than to the binary features or the whole-object (4-featured) conjunctions on which classifiers are trained to generate the FCI: would this spuriously bias FCI toward either positive or negative values? For the whole-object conjunction classifier, when voxels possess reliable information about just one of these mid-level conjunctions, the 16-way problem is reduced to a 4-way problem, facilitating above-chance classification and inflating the FCI toward more positive values. But, the feature classifiers would also benefit from mid-level conjunction-coding voxels: reliable information about one mid-level conjunction fully specifies 2 out of the 4 binary features when classifying any of the four stimuli that possess that mid-level conjunction, and provides a heuristic for classifying the other twelve stimuli that do not possess that mid-level conjunction (namely: "predict the feature values *not* included in the mid-level conjunction") that is neutral for eight of those stimuli and helpful for the other four. On average across all stimuli, feature classification performance would be boosted by mid-level conjunction-coding voxels, reducing the FCI toward more negative values. The net effect, across both classifier types and all stimuli, is that mid-level conjunction-sensitive voxels in an ROI should boost feature- and conjunction-classifier performance equally, pushing any non-zero FCI, either negative or positive, toward zero.

Multiple possible routes to a hybrid code

We suggested earlier that zero FCI (in the presence of above-chance classifier accuracy) might be produced by hybrid-coded activation patterns – a mixture of feature-sensitive and whole-object-sensitive voxels. But, given the presence of mid-level conjunctions in the stimuli, it is equally possible that zero FCI is produced by voxels sensitive to those mid-level conjunctions. However, this alternative interpretation of a zero FCI is in keeping with the spirit of the measure: whether a cortical region exhibits zero FCI because of a heterogeneous mix of feature- and conjunction-coding voxels or because of mid-level conjunction coding, that coding scheme should be considered intermediate, lying between strongly feature-coded and strongly whole-object-coded representations. In line with the idea that sensitivity to mid-level conjunctions tends to yield FCIs close to zero, whole brain analyses revealed that the zero-point in FCI lies near to area LOC – known to represent global shape (Drucker & Aguirre, 2009; Malach et al., 1995; Ostwald et al., 2008) – and area V4 – known to exhibit texture-selective responses (Arcizet, Jouffrais, & Girard, 2008; Dumoulin & Hess, 2007; Hanazawa & Komatsu, 2001). Global shape and texture constitute mid-level conjunctions in the present stimulus set.

A further possibility is that a hybrid code, comprising a mixture of feature- and conjunction-sensitive responses, is generated by not by combining spatially adjacent voxels of each type within an ROI, but rather by combining temporally adjacent responses of each type within the same neurons. Connor and colleagues (Brincat & Connor, 2006; Yau et al., 2013) recorded activation in monkey visual cortex and reported a dynamic neural code that evolved over time. Immediately after stimulus onset, neurons in posterior IT (pIT) and V4 were tuned to simple features but ~50-60ms later the same neurons responded non-linearly to multi-part configurations of those features (simple orientations were combined into multi-orientation curves in V4, and simple curved contour fragments into multi-part configurations in pIT). Because the BOLD signal has insufficient temporal resolution to distinguish between response types that evolve on the order of milliseconds, we cannot determine whether the hybrid code we observed was generated by a transition from feature- to conjunction-coding over time, but it is interesting that the areas producing zero FCI values in our data (occipito-temporal regions) lie near the regions identified in these electrophysiological studies (pIT and V4).

Changes in receptive field size with the emergence of conjunction-coding

The emergence of conjunction coding in the ventral and dorsal pathways is accompanied by a systematic change in another property – receptive field (RF) size. This is relevant to measuring feature and conjunction coding with our stimulus set, because the total stimulus area containing conjunction information is smaller than the area containing information about features. If larger RFs (which capture the whole stimulus) capture conjunction and feature information to a similar extent, whereas smaller RFs (which capture small stimulus subsections) capture relatively more feature information on average, does this bias small-RF brain regions toward feature-coding? Even with a small RF it is possible to capture conjunction information in this stimulus set, since each stimulus contains numerous points of intersection of the two outlines and fill patterns that uniquely define the stimulus. Moreover, even neurons with RFs in peripheral regions might code conjunction information, through collateral interactions or feedback from higher regions such that the feature information across a combination of neurons is pooled to construct the conjunction. So, in regions with small RFs, there is not a total absence of neurons that capture conjunction information, just a greater number of neurons able to capture feature information. Thus, the question of whether the greater spatial extent of feature than conjunction information in the stimuli biases regions with small RFs toward feature-coding rests upon whether the conjunction classifier in such a region is *impeded* by receiving inputs from a large number of voxels that code non-diagnostic information (features) along with inputs of diagnostic information (conjunctions). Empirically, it is not always the case that including less informative voxels in an ROI causes classifier performance to drop, relative to when a smaller subset of more informative voxels are used (e.g., Filimon, Rieth, Sereno, & Cottrell, 2015).

In addition, although increasing RF size could be described as a potential confound that shifts the balance of coding toward conjunctions in higher cortical areas, this ‘confound’ may in fact be part of the (adaptively-evolved) brain mechanism by which conjunction coding emerges in higher regions. In other words, part of the function of large receptive field sizes may be to enable conjunction-coding. Real-world objects at real-world viewing distances are most often captured by the visual system at sizes that afford apprehension of the whole object by an IT neuron, and apprehension of only small subsections of the object by a neuron in V1, and this is likely not an accident of natural selection. Thus, although an experimenter might seek to rule out any influence of RF size by presenting very small stimuli that are captured entirely by the RFs of V1 neurons, this scenario would not mimic natural viewing conditions. Moreover, the use of very small stimuli would introduce other systematic biases: small, centrally presented stimuli would be captured in V1 by only a small set neurons with foveally located RFs, but by almost all neurons in IT, so that the recorded activation data would exclude any contribution from peripheral RFs in early brain regions but not in later regions; and small stimuli might be rendered at such low resolution that the fine details of the conjunction are not perceived and stimuli are instead encoded as feature-like approximations in all brain regions.

Generalization to coding for other visual features and their conjunctions

We examined conjunctions of spatial frequency and contour. Spatial frequency is represented in V1 (Foster et al., 1985) whereas shape contour is represented in higher regions such as V4, pIT and LOC (Brincat & Connor, 2004, 2006; Drucker & Aguirre, 2009; Yau et al., 2013). However, classification of our binary contour features was good in V1 and V2 (Table 2) implying that, given the small receptive field sizes, these regions represented the contour features in terms of small, local oriented segments. To what extent do these findings generalize to conjunctions of other features such as orientation, color and motion? Orientation, like spatial frequency, is an elemental building block of vision, well-represented in V1 (Hubel & Wiesel, 1968). We suggest that shape is a visual attribute somewhat analogous to color and motion: for all three attributes, some information supporting classification is available in V1, but more holistic or categorical information emerges in higher regions – V4 for color (Brouwer & Heeger, 2013) and V3 or V5/MT for motion (Gegenfurtner et al., 1997; Movshon, Adelson, Gizzi, & Newsome, 1985). Thus, the feature conjunctions we examined may be typical of the part-to-whole integration operations putatively performed across the visual hierarchy. In line with this, empirical results for other conjunction types are relatively consistent with the present findings: conjunction-coding for 2-featured conjunctions (of complexity similar to our FMI) was reported in V2, V3 and parietal sites (Anzai et al., 2007; Baumgartner et al., 2013; Gegenfurtner et al., 1997; Pollmann et al., 2014). These prior studies did not examine feature-coding (as we define it), nor the evidence for a transition from feature- to conjunction-coding. However, we tentatively suggest that the posterior-to-anterior transition we observed is likely to apply for conjunctions of many flavors, because electrophysiology studies point to this conclusion (Kobatake & Tanaka, 1994; Rust & Dicarlo, 2010, 2012). An important caveat is that we do not claim that specific, quantitative aspects of our results can be generalized, such as the cortical locations of the transition point and of conjunction-coding sites. Quantitative details are likely specific to each stimulus set, and the present study merely provides a tool for their investigation.

Conclusions

Our novel method permits the systematic investigation of feature- and conjunction-coding, and may be applicable not just to vision but to other modalities such as audition or motor action (Lee, Turkeltaub, Granger, & Raizada, 2012; Wurm & Lingnau, 2015). Within vision, the method will enable future investigation of a range of features not included in our stimulus set (including color, orientation and motion), in order to examine how conjunction-coding emerges for different feature types and combinations. The present finding of a transition from feature- to conjunction- coding along both ventral and dorsal visual pathways has implications for theories of the functional architecture of visual object processing.

Acknowledgements

We thank David Huber for helpful discussions. This work was funded by NSF CAREER Award 1554871to R.A.C., by R01EY025872 to J.T.S. and by NSF grant BCS-0843773.

Tables

	Scan Session 1	**Scan Session 2**
Mean Proportion Correct (±.SEM)	0.903 ±.0185	0.947 ±.00894
Mean Arcsine(Proportion Correct) (±.SEM)	1.27 ±.0350	1.35 ±.0214

Table 1. Behavioral Performance in the Scanner. Mean accuracy on the 1-back repetition detection task performed in the scanner, with standard error of the mean (±SEM) across subjects. A t-test comparing the arcsine-transformed proportion correct scores indicated that performance was reliably better in Scan Session 2 (p = 0.0138).

A. Ventral Stream	**V1**	**V2v**	**V3v**	**LOC**
Feat 1 (SF1)	**0.791** ± 0.0261	**0.722** ± 0.0254	**0.686** ± 0.0178	**0.516** ± 0.0051
Feat 2 (Outline1)	**0.864** ± 0.0163	**0.675** ± 0.0188	**0.609** ± 0.0186	**0.550** ± 0.0101
Feat 3 (SF2)	**0.773** ± 0.0235	**0.679** ± 0.0204	**0.658** ± 0.0178	**0.526** ± 0.0092
Feat 4 (Outline2)	**0.853** ± 0.0160	**0.678** ± 0.0223	**0.657** ± 0.0176	**0.557** ± 0.0062
Overall Mean Acc.	**0.820** ± 0.0186	**0.688** ±0.0186	**0.653** ±0.0121	**0.537** ±0.00266
95% CI around overall mean	[0.815, 0.825]	[0.682,0.695]	[0.646, 0.659]	[0.530, 0.544]
B. Dorsal Stream		**V2d**	**V3d**	
Feat 1 (SF1)	-	**0.689** ± 0.0237	**0.652** ± 0.0195	-
Feat 2 (Outline1)	-	**0.777** ± 0.0239	**0.618** ± 0.0203	-
Feat 3 (SF2)	-	**0.685** ± 0.0211	**0.653** ± 0.0206	-
Feat 4 (Outline2)	-	**0.736** ± 0.0240	**0.692** ± 0.0280	-
Overall Mean Acc.	-	**0.722** ±0.02	**0.653** ±0.0146	-
95% CI around overall mean	-	[0.716, 0.728]	[0.647, 0.66]	-

Table 2. Feature-level Classifier Accuracy for Each Feature Type and Mean across All Feature Types. In both Ventral (A) and Dorsal (B) streams, rows 1-4 show the mean across subjects of the feature-level classifier accuracy for each feature type ± standard error of the mean (±SEM) across subjects. Chance = 0.5. SF = spatial frequency. Row 5 shows Overall Mean Accuracy across all four feature-level classifiers, ±SEM across subjects. Row 6 shows 95% CIs around the Overall Mean Accuracy determined by within-subjects bootstrap resampling with replacement over 10,000 iterations. All data were averaged over two sessions for each subject. Overall Mean Accuracy scores for each session in each subject were transformed into a log likelihood ratio (log odds; chance = 0) and submitted to a two-way repeated measures ANOVA, with factors Scan Session (1, 2) and ROI

(V1, V2v, V3v, V2d, V3d, LOC), revealing a main effect of ROI ($F[5,35]=83.14$, $p<0.001$, $\eta^2 = .922$), no significant effect of Scan Session ($F[1,7]=0.860$, $p=.39$, $\eta^2 = .109$), and a non-significant interaction ($F[5,35]=2.367$, $p=0.06$, $\eta^2 = .253$). Accuracy was lowest in region LOC, but a one-sample t-test revealed that the log odds of LOC accuracy (collapsed over sessions) exceeded chance performance ($M=0.1486$, $SD = 0.0303$, $t[7]=13.89$, $p<0.001$).

A. Ventral Stream	**V1**	**V2v**	**V3v**	**LOC**
Mean Accuracy	0.292 ± 0.0272	0.200 ± 0.0163	0.154 ± 0.0101	0.0926 ± 0.00701
95% CI	[0.280, 0.304]	[0.189, 0.210]	[0.144, 0.164]	[0.0848, 0.101]
D. Dorsal Stream		**V2d**	**V3d**	
Mean Accuracy	-	0.204 ± 0.0208	0.155 ± 0.0131	-
95% CI	-	[0.193, 0.215]	[0.145, 0.164]	-

Table 3. Object-level Classifier Accuracy. Mean accuracy of object-level classifier by ROI (average over two sessions for each subject), with standard error of the mean across subjects (±SEM) and 95% Confidence Intervals (CIs) around the mean, determined by within-subjects bootstrap resampling with replacement over 10,000 iterations. Chance = 0.0625. Accuracy scores for each session in each subject were transformed into a log likelihood ratio (log odds; chance = -2.71) and submitted to a two-way repeated measures ANOVA, with factors Scan Session (1, 2) and ROI (V1, V2v, V3v, V2d, V3d, LOC), revealing a main effect of ROI ($F[5,35]=61.96$, $p<0.001$, $\eta^2 = .90$), no significant effect of Scan Session ($F[1,7]=0.312$, $p=.59$, $\eta^2 = .043$), and no significant interaction ($F[5,35]=1.61$, $p=0.183$, $\eta^2 = .187$). Accuracy was lowest in region LOC, but a one-sample t-test revealed that the log odds of LOC accuracy (collapsed over sessions) exceeded chance performance ($M=-2.30$, $SD = 0.2321$, $t[7]=4.95$, $p<0.001$).

A. Ventral Stream	**V1**	**V2v**	**V3v**	**LOC**
Subject 1	-0.564	-0.081	-0.308	0.080
Subject 2	-0.467	-0.271	-0.137	0.188
Subject 3	-0.425	-0.251	-0.110	0.366
Subject 4	-0.476	-0.161	-0.284	0.391
Subject 5	-0.453	-0.301	-0.319	0.102
Subject 6	-0.464	-0.151	-0.148	-0.145
Subject 7	-0.584	-0.304	-0.209	-0.091
Subject 8	-0.494	0.089	-0.133	-0.181
Mean FCI	**-0.491**	**-0.179**	**-0.206**	**0.089**
SEM	0.020	0.048	0.030	0.078
95% CI	[-0.545, -0.442]	[-0.252, -0.110]	[-0.292, -0.125]	[-0.0558, 0.192]
B. Dorsal Stream		**V2d**	**V3d**	
Subject 1	-	-0.207	-0.088	-
Subject 2	-	-0.207	-0.085	-
Subject 3	-	-0.298	-0.284	-
Subject 4	-	-0.440	-0.313	-
Subject 5	-	-0.497	-0.222	-
Subject 6	-	-0.396	-0.276	-
Subject 7	-	-0.345	0.055	-
Subject 8	-	-0.418	-0.431	-
Mean FCI	**-**	**-0.351**	**-0.206**	**-**
SEM	-	0.038	0.055	-
95% CI	-	[-0.422,-0.284]	[-0.293, -0.122]	-

Table 4. Feature-Conjunction Index in Visual Cortical ROIs. Individual subjects' FCI by ROI (average over two sessions for each subject), with mean, standard error of the mean (±SEM) across subjects, and 95% CIs determined by within-subjects bootstrap sampling with replacement over 10,000 iterations (see *Methods*). Mean values are also shown graphically in Figure 3.

	Accuracy on Final Session	**RT-Set Size Slope First Session (ms/item)**	**RT-Set Size Slope Final Session (ms/item)**
Mean (±.SEM)	.893 ±.020	297 ± 37	200 ±20
SEM	±.020	±37	

Table 5. Behavioral Performance During Training on the Conjunctive Visual Search Task. Accuracy on the final session (and in all earlier sessions, not reported) far exceeded the chance performance level of 50%. RT-set size slopes decreased significantly from the first to the final session (p<.05) but were still significantly greater than 0 in the final session (p<0.0001). Shiffrin and Lightfoot (1997) trained subjects for many more sessions of conjunctive visual search (~50 days as compared to 11.1 days in the present study) and observed RT-set size slopes as low as 50ms/item on the final session.

A. Ventral Stream	**V1**	**V2v**	**V3v**	**LOC**
Mid-level Conj. 1 (Global Shape)	**0.689** ± 0.029	**0.473** ±0.021	**0.400** ±0.012	**0.327** ± 0.010
Mid-level Conj. 2 (Texture)	**0.610** ± 0.036	**0. 517** ± 0.026	**0.471** ±0.023	**0.281** ± 0.007
Mid-level Conj. 3 (Right Component)	**0.666** ± 0.032	**0.493** ± 0.021	**0.413** ± 0.017	**0.296** ± 0.007
Mid-level Conj. 4 (Left Component)	**0.645** ± 0.033	**0.463** ± 0.025	**0.433** ± 0.015	**0.293** ± 0.006
Overall Mean Acc.	**0.653** ± 0.031	**0.487** ± 0.022	**0.429** ±0.015	**0.299** ±0.005
B. Dorsal Stream		**V2d**	**V3d**	
Mid-level Conj. 1 (Global Shape)	-	**0.573** ±0.025	**0.494** ± 0.017	-
Mid-level Conj. 2 (Texture)	-	**0.491** ± 0.031	**0.453** ± 0.023	-
Mid-level Conj. 3 (Right Component)	-	**0.562** ± 0.028	**0.469** ± 0.027	-
Mid-level Conj. 4 (Left Component)	-	**0.504** ± 0.028	**0.475** ± 0.029	-
Overall Mean Acc.	-	**0.533** ± 0.026	**0.473** ± 0.021	-

Table 6. Mid-Level Conjunction Classifier Accuracies. Accuracy of the mid-level conjunction classifier, by ROI, with mean and standard error of the mean (±SEM) across subjects. Data were averaged over two sessions for each subject. Chance = 0.25. Overall Mean Accuracy scores for each session in each subject were transformed into a log likelihood ratio (log odds; chance = -0.477) and

submitted to a two-way repeated measures ANOVA, with factors Scan Session (1, 2) and ROI (V1, V2v, V3v, V2d, V3d, LOC), revealing a main effect of ROI ($F[5,35]=76.65$, $p<0.001$, $\eta^2 = .92$), no significant effect of Scan Session ($F[1,7]=0.939$, $p=.37$, $\eta^2 = .12$), and no significant interaction ($F[2.272,35]=3.12$, $p=0.067$, $\eta^2 = .31$; Greenhouse-Geisser correction for violation of sphericity). Accuracy was lowest in region LOC, but a one-sample t-test revealed that the log odds of LOC accuracy (collapsed over sessions) exceeded chance ($M=-0.370$, $SD = 0.010$, $t[7]=10.68$, $p<0.001$). FMI and MCI values derived using these classifier accuracies (in combination with feature- and whole object-classifier accuracies, respectively) are shown in Figures 8 and 9, respectively. Note that before deriving FMI and MCI using classifier accuracies, we removed any session-subject-ROI instances for which no classifier exceeded chance, as determined by a binomial test, see Methods.

A. Ventral Stream	**V1**	**V2v**	**V3v**	**LOC**
Mid-level Conj. 1 (Global Shape)	**0.0005** ± 0.026	**-0.0419** ± 0.039	**-0.044** ± 0.035	**0.181** ± 0.038
Mid-level Conj. 2 (Texture)	**-0.089** ± 0.024	**0.117** ± 0.028	**0.086** ± 0.035	**-0.095** ± 0.036
Mid-level Conj. 3 (Right Component)	**0.076** ± 0.017	**0.0029** ± 0.028	**-0.108** ± 0.024	**0.116** ± 0.063
Mid-level Conj. 4 (Left Component)	**-0.148** ± 0.026	**-0.0016** ± 0.024	**0.061** ± 0.024	**-0.044** ± 0.053
Overall Mean FMI	**-0.040** ± 0.005	**0.019** ± 0.011	**-0.001** ± 0.008	**0.041** ± 0.023
B. Dorsal Stream		**V2d**	**V3d**	
Mid-level Conj. 1 (Global Shape)	-	**0.056** ±0.031	**0.195** ± 0.060	-
Mid-level Conj. 2 (Texture)	-	**-0.037** ±0.027	**-0.003** ± 0.048	-
Mid-level Conj. 3 (Right Component)	-	**0.157** ±0.023	**0.090** ± 0.035	-
Mid-level Conj. 4 (Left Component)	-	**-0.134** ±0.021	**0.110** ± 0.049	-
Overall Mean FMI	-	**0.011** ±0.012	**0.098** ± 0.016	-

Table 7. Feature Mid-Level Conjunction Index (FMI) for the Four Plausible Mid-level Conjunctions, Separately. FMI for each mid-level conjunction separately, by ROI, with mean and standard error of the mean (±SEM) across subjects. Mean values over all 4 mid-level conjunctions for each ROI are also shown in Figure 8. Data were averaged over two sessions for each subject.

Figure Legends

Figure 1. Stimulus Construction, Task Protocol and Multivariate Pattern Classifiers

A. Stimulus Construction. The four binary features from which the 16 object-level stimuli are composed (full stimulus set shown in panel D). Each feature has two possible values: A and B.

B. Task Protocol. In each scan session, participants completed 10 experimental runs, each lasting 264 seconds (44 trials of 6 sec duration). A run contained 34-36 stimulus trials (two presentations each of the 16 stimuli in the set, ordered pseudo-randomly, in addition to 2-4 stimuli chosen pseudo-randomly from the set and inserted to create immediate repeats) and 8-10 nulls trials. On stimulus presentation trials, the fixation point was red, stimulus duration was 3sec and the inter-stimulus interval varied between 2.5 and 3.5sec; participants performed a 1-back repetition detection task. On null trials, the fixation point changed to green; participants indicated by button press when they detected a slight dimming of the fixation point, which occurred once or twice per null trial. Participants also completed several sessions of visual search training between the two scans, but we detected no effect of training on our measure of feature- and conjunction-coding (FCI) in cortex.

C. Feature Classification. Four separate feature classifiers were trained, one for each binary feature defined in the stimulus set. The four feature classification problems are shown in the four panels (Right Spatial Frequency, Right Outline, Left Spatial Frequency, Left Outline) in which the designation of stimuli to feature categories is indicated with red and green boxes. Classifiers used a support vector machine trained with hold-one-out cross-validation.

D. Object Classification. A single object-level classifier was trained to classify the stimuli into 16 categories, each corresponding to a unique stimulus.

E. Calculation of the Feature Conjunction Index (FCI). The product of the four feature-level accuracies was used to predict – independently for each trial – the accuracy of a hypothetical object-level classifier whose performance depends only on feature-level information. On each trial, the four feature-classifier responses (defined as 0 or 1 for incorrect or correct) were multiplied to produce a value of 0 or 1 (incorrect or correct) for the hypothetical object-level classifier. Next, the empirically observed object-level classifier accuracy (derived from the sixteen-way conjunction classifier) and the hypothetical object-level accuracy (predicted from the four feature classifiers) were averaged over trials and placed in a log ratio (Equation 1). When the empirically observed object classifier accuracy exceeds the hypothetical object accuracy predicted from feature classifier accuracies, FCI is positive; when the feature classifier accuracies predict better object-level knowledge accuracy than is obtained by the object classifier, FCI is negative (see Figure 2).

Figure 2. Classifier Accuracy and FCI for Synthetic Data

Top and Bottom panels show simulation results for synthetic data generated with Feature-Coded and Conjunction-Coded activation pattern templates, respectively. Results shown are from Method 1 of generating synthetic data; Method 2 results are not shown but were very similar, and produced the same conclusions from statistical tests. Error bars show standard error of the mean (SEM) for FCI. Gray box for each dataset shows the range of SNR values that produced mean classifier accuracies falling within the range observed in ROI-based analyses of the empirical data (see Tables 2 and 3). We focus upon FCI values within the gray box as being representative of plausible outcomes from empirical BOLD data for each underlying template. The lower bound of each gray box is set to exclude from the box all SNR values at which neither the Feature nor the 16-way Conjunction Classifier Accuracy exceeded the lowest accuracy observed in the empirical ROIs (0.537 for Feature Classification; 0.0926 for Conjunction Classification; see Tables 2 and 3). The upper bound of each gray box is set to exclude all SNR values at which either the Feature or the Conjunction Classifier Accuracy exceeded the maximum accuracy observed in classifiers trained on empirical ROI-based data (0.864 for Feature Classification; 0.292 for Conjunction Classification; see

Tables 2 and 3). The extremely high FCI values produced by synthetic conjunction-coded data were never observed in the empirical data; this may be due in part to the fact that the high SNR required to produce high positive FCI values in the synthetic data does not exist in cortical regions that exhibit conjunction coding (i.e., regions exhibiting positive FCI values yield lower accuracy, presumably because of lower SNR).

Figure 3. Feature Conjunction Indices (FCI) Derived from ROI-based Analyses

Mean FCI (averaged over two sessions in each subject) for ROIs in early ventral visual stream (top panel: V1, V2v, V3v, LOC) and early dorsal stream (bottom panel: V1, V2d V3d). V1 is duplicated in top and bottom plots for ease of comparison. FCI is the natural logarithm of the ratio of object classifier accuracy to the product of the four feature classifier accuracies (see Figure 1 and *Methods*). Positive FCI reflects conjunction-coding; negative FCI reflects feature-coding. Gray bars show group mean; plotted points show individual subjects, where each unique marker corresponds to the same individual subject across ROIs. See Table 4 for bootstrapped 95% Confidence Intervals (CIs) around the means. n = 8 subjects, 2 scan sessions each.

Figure 4: FCI Derived from Whole-brain Searchlight Analyses

Group mean FCI produced by a searchlight MVPA analysis assessing conjunction- versus feature-coding throughout visual cortex. A sphere of radius 5 functional voxels was swept through the imaged volume, constrained by a subject-specific grey-matter mask encompassing occipital, temporal and posterior parietal cortex. Taking each voxel in turn as the centroid of a spherical ROI, the feature and object classifiers were trained and their accuracies combined to produce a FCI which was entered into the map at the location of the centroid voxel. Orange indicates positive FCI (conjunction-coding), blue indicates negative FCI (feature-coding). Centroid voxels for which classifier performance did not exceed chance, as determined by a binomial test, were removed from individual subject maps (see *Methods*). FCI maps were constructed for each subject and scan session individually (see Figure 5), then spatially smoothed using a Gaussian kernel (FWHM=2 functional voxels). Smoothed maps were averaged across two sessions for each subject, and across subjects. Scale is truncated at ±0.25 for optimal visualization of the data; some voxels possess FCI values > +0.25 or < -0.25.

Figure 5. Individual Subject FCI Maps.

Raw FCI values plotted separately for the two scans (1, 2) in three individual subjects (G, H and N). The two leftmost columns correspond to the first scan, and the two rightmost columns correspond to the second scan; each row depicts an individual subject. Color of a voxel indicates FCI for the spherical ROI surrounding it. Absence of color indicates the voxel was screened out because either (1) it was not included in the anatomical (grey matter) mask, (2) the sphere surrounding the voxel contained too few voxels, or (3) accuracy for none of the classifiers exceeded threshold. There are many more absent voxels here than in the group mean FCI map (Figure 4) because the spatial smoothing and averaging used in creating the group mean map eliminated many absent voxels. Scale is truncated at ±0.5 for optimal visualization; some voxels possess FCI values > +0.5 or < -0.5. The scale differs from that of Figure 4 because data within individual maps span a greater range.

Figure 6. Quantification of the transition from feature- to conjunction-coding in the ventral and dorsal streams.

Insets show, in pink, the approximate extent and position of the three defined vectors Posterior Ventral, Anterior Ventral and Dorsal; in blue, the projection of a voxel location onto the vector to derive a scalar value for the voxel position; in green, the bounding box defining the brain region included as part of each pathway (a subject-specific anatomical mask including only grey matter in occipital, temporal and parietal lobes was also applied). Plots show the best fitting regression lines relating the location of a voxel in each of the three pathways to the FCI for the spherical ROI surrounding the voxel. Each line shows one subject in one hemisphere; colors indicate different subjects; solid and dashed lines show Left and Right hemispheres, respectively. The far endpoint of the vector is more distant from occipital cortex (i.e., the vector is longer) for the Dorsal than the Posterior Ventral pathway: this may account for the higher FCI value at the vector endpoint in the Dorsal than the Posterior Ventral pathway, given that regression line slopes in the Dorsal and Posterior Ventral pathways were similar (x and y-axes use the same scale for all 3 plots).

Figure 7. Cortical sites of feature- and conjunction-coding observed at the group level.

Statistical map shows the results of a t-test at each voxel comparing the group mean FCI value associated with the spherical ROI surrounding that voxel to zero. The map was thresholded at p=0.05, two-tailed (FDR-corrected for multiple comparisons). Blue voxels possess FCI values significantly less than zero (feature-coding); orange voxels possess FCI values significantly greater than zero (conjunction-coding). All but one voxels with statistically reliable negative FCI values were located in occipital cortex, whereas all voxels with statistically reliable positive FCI values were located anterior or superior to the occipital feature-coding regions. Voxels exhibiting significant conjunction coding appeared in multiple sites bilaterally, including posterior parietal lobe, fusiform gyrus, parahippocampal gyrus (including left perirhinal cortex), hippocampus and the anterior temporal pole. Axial slices are radiologically flipped (left hemisphere appears on the right). In sagittal slices, positive X co-ordinates indicate right hemisphere.

Figure 8. Feature vs. Mid-level-Conjunction Indices (FMI) Derived from ROI-based Analyses

Mean FMI (averaged over two sessions in each subject) for ROIs in early ventral visual stream (top panel: V1, V2v, V3v, LOC) and early dorsal stream (bottom panel: V1, V2d V3d). V1 is duplicated in top and bottom plots for ease of comparison. FMI is the natural logarithm of the ratio of the mid-level conjunction classifier accuracy to the product of the two feature classifier accuracies for the features defining that mid-level conjunction (see *Methods*). Positive FMI reflects a preference for coding mid-level conjunctions over features; negative FMI reflects the reverse preference. Gray bars show group mean; plotted points show individual subjects, where each unique marker corresponds to the same individual subject across ROIs.

Figure 9. Mid-Level-Conjunction vs. Whole-Object-Conjunction Indices (MCI) Derived from ROI-based Analyses

Mean MCI (averaged over two sessions in each subject) for ROIs in early ventral visual stream (top panel: V1, V2v, V3v, LOC) and early dorsal stream (bottom panel: V1, V2d V3d). V1 is duplicated in top and bottom plots for ease of comparison. MCI is the natural logarithm of the ratio of the whole-object classifier accuracy to the product of the accuracies of each pair of mid-level conjunction classifiers that define the whole object (see *Methods*). Positive MCI reflects a preference for coding whole objects over mid-level conjunctions; negative MCI reflects the reverse preference. Gray bars show group mean; plotted points show individual subjects, where each unique marker corresponds to the same individual subject across ROIs.

References

Anzai, A., Peng, X., & Van Essen, D. C. (2007). Neurons in monkey visual area V2 encode combinations of orientations. *Nat Neurosci, 10*(10), 1313-1321.

Arcizet, F., Jouffrais, C., & Girard, P. (2008). Natural textures classification in area V4 of the macaque monkey. *Exp Brain Res, 189*(1), 109-120.

Ashbridge, E., Walsh, V., & Cowey, A. (1997). Temporal aspects of visual search studied by transcranial magnetic stimulation. *Neuropsychologia, 35*(8), 1121-1131.

Baeck, A., Wagemans, J., & Op de Beeck, H. P. (2013). The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: the weighted average as a general rule. *Neuroimage, 70*, 37-47.

Baumgartner, F., Hanke, M., Geringswald, F., Zinke, W., Speck, O., & Pollmann, S. (2013). Evidence for feature binding in the superior parietal lobule. *Neuroimage, 68*, 173-180.

Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat Neurosci, 7*(8), 880-886.

Brincat, S. L., & Connor, C. E. (2006). Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron, 49*(1), 17-24.

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *Journal of Neuroscience, 29*(44), 13992-14003.

Brouwer, G. J., & Heeger, D. J. (2013). Categorical Clustering of the Neural Representation of Color. *Journal of Neuroscience, 33*(39), 15454-15465.

Bussey, T. J., & Saksida, L. M. (2002). The organization of visual object representations: a connectionist model of effects of lesions in perirhinal cortex. *European Journal of Neuroscience, 15*(2), 355-364.

Chang, C. C., & Lin, C. J. (2011). LIBSVM: A Library for Support Vector Machines. *Acm Transactions on Intelligent Systems and Technology, 2*(3).

Cohen, A., & Rafal, R. D. (1991). Attention and Feature Integration: Illusory Conjunctions in a Patient With a Parietal Lobe Lesion. *Psychological Science, 2*(2), 106-110.

Cowell, R. A., Bussey, T. J., & Saksida, L. M. (2010). Functional Dissociations within the Ventral Object Processing Pathway: Cognitive Modules or a Hierarchical Continuum? *Journal of Cognitive Neuroscience, 22*(11), 2460-2479.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-Selective Properties of Inferior Temporal Neurons in the Macaque. *Journal of Neuroscience, 4*(8), 2051-2062.

Drucker, D. M., & Aguirre, G. K. (2009). Different Spatial Scales of Shape Similarity Representation in Lateral and Ventral LOC. *Cerebral Cortex, 19*(10), 2269-2280.

Dumoulin, S. O., & Hess, R. F. (2007). Cortical specialization for concentric shape processing. *Vision Res, 47*(12), 1608-1613.

Eckhorn, R. (1999a). Neural mechanisms of scene segmentation: recordings from the visual cortex suggest basic circuits for linking field models. *IEEE Trans Neural Netw, 10*(3), 464-479.

Eckhorn, R. (1999b). Neural mechanisms of visual feature binding investigated with microelectrodes and models. *Visual Cognition, 6*(3-4), 231-265.

Engel, S. A. (2005). Adaptation of oriented and unoriented color-selective neurons in human visual areas. *Neuron, 45*(4), 613-623.

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E. J., et al. (1994). fMRI of human visual cortex. *Nature, 369*(6481), 525.

Epstein, R. A., Higgins, J. S., Parker, W., Aguirre, G. K., & Cooperman, S. (2006). Cortical correlates of face and scene inversion: a comparison. *Neuropsychologia, 44*(7), 1145-1158.

Erez, J., Cusack, R., Kendall, W., & Barense, M. D. (2015). Conjunctive Coding of Complex Object Features. *Cerebral Cortex*.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex, 1*(1), 1-47.

Filimon, F., Rieth, C. A., Sereno, M. I., & Cottrell, G. W. (2015). Observed, Executed, and Imagined Action Representations can be Decoded From Ventral and Dorsal Areas. *Cereb Cortex, 25*(9), 3144-3158.

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cereb Cortex, 23*(4), 814-823.

Foster, K. H., Gaska, J. P., Nagler, M., & Pollen, D. A. (1985). Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *J Physiol, 365*, 331-363.

Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. *Science, 269*(5225), 853-855.

Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for Visual Features of Objects in Monkey Inferotemporal Cortex. *Nature, 360*(6402), 343-346.

Fukushima, K. (1980). Neocognitron - a Self-Organizing Neural Network Model for a Mechanism of Pattern-Recognition Unaffected by Shift in Position. *Biological Cybernetics, 36*(4), 193-202.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat Neurosci, 2*(6), 568-573.

Gegenfurtner, K. R., Kiper, D. C., & Levitt, J. B. (1997). Functional properties of neurons in macaque area V3. *J Neurophysiol, 77*(4), 1906-1923.

Goda, N., Tachibana, A., Okazawa, G., & Komatsu, H. (2014). Representation of the material properties of objects in the visual cortex of nonhuman primates. *J Neurosci, 34*(7), 2660-2673.

Grill-Spector, K. (2003). The neural basis of object perception. *Curr Opin Neurobiol, 13*(2), 159-166.

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Human Brain Mapping, 6*(4), 316-328.

Hanazawa, A., & Komatsu, H. (2001). Influence of the direction of elemental luminance gradients on the responses of V4 cells to textured surfaces. *J Neurosci, 21*(12), 4490-4497.

Henriksson, L., Nurminen, L., Hyvarinen, A., & Vanni, S. (2008). Spatial frequency tuning in human retinotopic visual areas. *J Vis, 8*(10), 5 1-13.

Hiramatsu, C., Goda, N., & Komatsu, H. (2011). Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. *Neuroimage, 57*(2), 482-494.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience, 3*(3), 284-291.

Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *Ieee Transactions on Neural Networks, 13*(2), 415-425.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat. *Journal of Neurophysiology, 28*, 229-289.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive Fields and Functional Architecture of Monkey Striate Cortex. *Journal of Physiology-London, 195*(1), 215-&.

Humphreys, G. W., Cinel, C., Wolfe, J., Olson, A., & Klempen, N. (2000). Fractionating the binding process: neuropsychological evidence distinguishing binding of form from binding of surface features. *Vision Res, 40*(10-12), 1569-1596.

Jeong, S. K., & Xu, Y. (2016). Behaviorally Relevant Abstract Object Identity Representation in the Human Parietal Cortex. *J Neurosci, 36*(5), 1607-1619.

Johnson, E. N., Hawken, M. J., & Shapley, R. (2008). The orientation selectivity of color-responsive neurons in macaque V1. *J Neurosci, 28*(32), 8096-8106.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8*(5), 679-685.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302-4311.

Kobatake, E., & Tanaka, K. (1994). Neuronal Selectivities to Complex Object Features in the Ventral Visual Pathway of the Macaque Cerebral-Cortex. *Journal of Neurophysiology, 71*(3), 856-867.

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nat Neurosci, 11*(2), 224-231.

Kreiman, G., Hung, C. P., Kraskov, A., Quiroga, R. Q., Poggio, T., & DiCarlo, J. J. (2006). Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. *Neuron, 49*(3), 433-445.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America, 103*(10), 3863-3868.

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*(6), 1126-1141.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp, 10*(3), 120-131.

Large, M. E., Aldcroft, A., & Vilis, T. (2005). Perceptual continuity and the emergence of perceptual persistence in the ventral visual pathway. *J Neurophysiol, 93*(6), 3453-3462.

Lee, Y. S., Turkeltaub, P., Granger, R., & Raizada, R. D. (2012). Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. *J Neurosci, 32*(11), 3942-3948.

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cereb Cortex, 11*(4), 287-297.

Macevoy, S. P., & Epstein, R. A. (2009). Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr Biol, 19*(11), 943-947.

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. *Proc Natl Acad Sci U S A, 92*(18), 8135-8139.

Mazer, J. A., Vinje, W. E., McDermott, J., Schiller, P. H., & Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proc Natl Acad Sci U S A, 99*(3), 1645-1650.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1985). The analysis of moving visual patterns. In C. Chagas, R. Gattass & C. G. Gross (Eds.), *Pattern Recognition Mechanisms: Pontificiae Academiae Scientiarum Scripta Varia* (Vol. 54, pp. 117–151). Rome: Vatican Press.

Muggleton, N. G., Cowey, A., & Walsh, V. (2008). The role of the angular gyrus in visual conjunction search investigated using signal detection analysis and transcranial magnetic stimulation. *Neuropsychologia, 46*(8), 2198-2202.

Ostwald, D., Lam, J. M., Li, S., & Kourtzi, Z. (2008). Neural coding of global form in the human visual cortex. *J Neurophysiol, 99*(5), 2456-2469.

Parks, N. A., Mazzi, C., Tapia, E., Savazzi, S., Fabiani, M., Gratton, G., et al. (2015). The influence of posterior parietal cortex on extrastriate visual activity: A concurrent TMS and fast optical imaging study. *Neuropsychologia, 78*, 153-158.

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *J Neurophysiol, 82*(5), 2490-2502.

Perrett, D. I., & Oram, M. W. (1993). Neurophysiology of Shape Processing. *Image and Vision Computing, 11*(6), 317-333.

Pollmann, S., Zinke, W., Baumgartner, F., Geringswald, F., & Hanke, M. (2014). The right temporo-parietal junction contributes to visual feature binding. *Neuroimage, 101*, 289-297.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019-1025.

Rowe, J. B., Stephan, K. E., Friston, K., Frackowiak, R. S. J., & Passingham, R. E. (2005). The prefrontal cortex shows context-specific changes in effective connectivity to motor or visual cortex during the selection of action or colour. *Cerebral Cortex, 15*(1), 85-95.

Rust, N. C., & Dicarlo, J. J. (2010). Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *J Neurosci, 30*(39), 12978-12995.

Rust, N. C., & DiCarlo, J. J. (2012). Balanced increases in selectivity and tolerance produce constant sparseness along the ventral visual stream. *J Neurosci, 32*(30), 10170-10182.

Serences, J. T., Saproo, S., Scolari, M., Ho, T., & Muftuler, L. T. (2009). Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. *Neuroimage, 44*(1), 223-231.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science, 268*(5212), 889-893.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci U S A, 104*(15), 6424-6429.

Seymour, K., Clifford, C. W., Logothetis, N. K., & Bartels, A. (2009). The coding of color, motion, and their conjunction in the human visual cortex. *Current Biology, 19*(3), 177-183.

Seymour, K., Clifford, C. W., Logothetis, N. K., & Bartels, A. (2010). Coding and binding of color and form in visual cortex. *Cerebral Cortex, 20*(8), 1946-1954.

Shiffrin, R., & Lightfoot, N. (1997). Perceptual Learning of Alphanumeric-Like Characters. In R. L. Goldstone, D. L. Medin & P. G. Schyns (Eds.), *Psychology of Learning and Motivation: Perceptual Learning* (Vol. 36, pp. 45-81): Elsevier.

Shigihara, Y., & Zeki, S. (2013). Parallelism in the brain's visual form system. *Eur J Neurosci, 38*(12), 3712-3720.

Shigihara, Y., & Zeki, S. (2014). Parallel processing in the brain's visual form system: an fMRI study. *Front Hum Neurosci, 8*, 506.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci, 18*, 555-586.

Song, J. H., & Jiang, Y. (2006). Visual working memory for simple and complex features: an fMRI study. *Neuroimage, 30*(3), 963-972.

Stark, C. E., & Squire, L. R. (2001). When zero is not zero: the problem of ambiguous baseline conditions in fMRI. *Proc Natl Acad Sci U S A, 98*(22), 12760-12766.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain : 3-dimensional proportional system : an approach to cerebral imaging*. New York, NY: Thieme Medical Publishers.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*, 109-139.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cogn Psychol, 12*(1), 97-136.

van den Honert, R. N., McCarthy, G., & Johnson, M. K. (2017). Holistic versus feature-based binding in the medial temporal lobe. *Cortex, 91*, 56-66.

von der Heydt, R., Peterhans, E., & Dursteler, M. R. (1992). Periodic-pattern-selective cells in monkey visual cortex. *J Neurosci, 12*(4), 1416-1434.

Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology, 51*(2), 167-194.

Wurm, M. F., & Lingnau, A. (2015). Decoding actions at different levels of abstraction. *J Neurosci, 35*(20), 7727-7735.

Xu, Y. (2009). Distinctive neural mechanisms supporting visual object individuation and identification. *J Cogn Neurosci, 21*(3), 511-518.

Yau, J. M., Pasupathy, A., Brincat, S. L., & Connor, C. E. (2013). Curvature processing dynamics in macaque area V4. *Cereb Cortex, 23*(1), 198-209.

Zhang, J., Liu, J., & Xu, Y. (2015). Neural decoding reveals impaired face configural processing in the right fusiform face area of individuals with developmental prosopagnosia. *J Neurosci, 35*(4), 1539-1548.

Zoccolan, D., Kouh, M., Poggio, T., & DiCarlo, J. J. (2007). Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *J Neurosci, 27*(45), 12292-12307.