

ORIGINAL ARTICLE

Conjunctive Coding of Complex Object Features

Jonathan Erez¹, Rhodri Cusack^{2,3}, William Kendall⁴, and Morgan D. Barense^{1,5}

¹Department of Psychology, University of Toronto, Toronto, ON, Canada M5S 3G3, ²Department of Psychology, University of Western Ontario, London, ON, Canada N6A 5C2, ³The Brain and Mind Institute, London, ON, Canada N6A 5B7, ⁴Department of Psychology, University of British Columbia, Vancouver, BC, Canada V6T 1Z4, and ⁵Rotman Research Institute, Toronto, ON, Canada M6A 2E1

Address correspondence to Jonathan Erez, Department of Psychology, University of Toronto, 100 St George Street, Toronto, ON, Canada M5S 3G3.
Email: jonathan.erez@utoronto.ca

Abstract

Critical to perceiving an object is the ability to bind its constituent features into a cohesive representation, yet the manner by which the visual system integrates object features to yield a unified percept remains unknown. Here, we present a novel application of multivoxel pattern analysis of neuroimaging data that allows a direct investigation of whether neural representations integrate object features into a whole that is different from the sum of its parts. We found that patterns of activity throughout the ventral visual stream (VVS), extending anteriorly into the perirhinal cortex (PRC), discriminated between the same features combined into different objects. Despite this sensitivity to the unique conjunctions of features comprising objects, activity in regions of the VVS, again extending into the PRC, was invariant to the viewpoints from which the conjunctions were presented. These results suggest that the manner in which our visual system processes complex objects depends on the explicit coding of the conjunctions of features comprising them.

Key words: feature integration, hierarchical object processing, MVPA, perirhinal cortex, view-invariance

Introduction

How objects are represented in the brain is a core issue in neuroscience. In order to coherently perceive even a single object, the visual system must integrate its features (e.g., shape, color) into a unified percept (sometimes called the “binding problem”) and recognize this object across different viewing angles, despite the drastic variability in appearance caused by shifting viewpoints (the “invariance problem”). These are the most computationally demanding challenges faced by the visual system, yet humans can perceive complex objects across different viewpoints with exceptional ease and speed (Thorpe et al. 1996). The mechanism underlying this feat is one of the central unsolved puzzles in cognitive neuroscience.

Two main classes of models have been proposed. The first are hierarchical models, in which representations of low-level features are transformed into more complex and invariant representations as information flows through successive stages of the ventral visual stream (VVS), a series of anatomically linked

cortical fields originating in V1 and extending into the temporal lobe (Hubel and Wiesel 1965; Desimone and Ungerleider 1989; Gross 1992; Tanaka 1996; Riesenhuber and Poggio 1999). These models assume *explicit conjunctive coding* of bound features: posterior VVS regions represent low-level features and anterior regions represent increasingly complex and invariant conjunctions of these simpler features. In contrast, an alternative possibility is a *non-local binding mechanism*, in which the perception of a unitized object does not necessitate explicit conjunctive coding of object features per se, but rather, the features are represented independently and bound by co-activation. Such a mechanism could include the synchronized activity of spatially distributed neurons that represent the individual object features (Singer and Gray 1995; Uhlhaas et al. 2009), or a separate brain region that temporarily reactivates and dynamically links otherwise disparate feature representations (Eckhorn 1999; Devlin and Price 2007). Thus, an explicit conjunctive coding mechanism predicts that the neural representation for a whole object should be different from the sum of its parts, whereas a nonlocal binding

mechanism predicts that the whole should not be different from the sum of the parts, because the unique conjunctive representations are never directly coded.

The neuroimaging method of multivoxel pattern analysis (MVPA) offers promise for making more subtle distinctions between representational content than previously possible (Haxby et al. 2001; Kamitani and Tong 2005). Here, we used a novel variant of MVPA to adjudicate between these 2 mechanisms. Specifically, we measured whether the representation of a whole object differed from the combined representations of its constituent features (i.e., explicit conjunctive coding), and whether any such conjunctive representation was view-invariant. We examined the patterns of neural activity evoked by 3 features distributed across 2 individually presented objects during a 1-back task (Fig. 1). Our critical contrast measured the additivity of patterns evoked by different conjunctions of features across object pairs: $A + BC$ versus $B + AC$ versus $C + AB$, where A, B, and C each represent an object comprising a single feature, and AB, BC, and AC each represent an object comprising conjunctions of those features (Fig. 2A,B). Importantly, in this “conjunction contrast,” the object pairs were identical at the feature level (all contained A, B, and C), but differed in their conjunction (AB vs. BC vs. AC), allowing a clean assessment of the representation pertaining to the conjunction, over and above any information regarding the component features. This balanced design also ensured that mnemonic demands were matched across comparisons. A finding of equivalent additivity (i.e., if $A + BC = B + AC = C + AB$) would indicate that information pertaining to the specific conjunctions is not represented in the patterns of activity—consistent with a nonlocal binding mechanism in which the features comprising an object are bound by their co-activation. In contrast, if the pattern sums are not equivalent (i.e., if $A + BC \neq B + AC \neq C + AB$), then the neural code must be conjunctive, representing information about the specific conjunctions of features over and above information pertaining to the individual features themselves—consistent with an explicit conjunctive coding mechanism.

An important potential benefit of explicit conjunctive coding of whole objects is to provide stability of representation across changes in viewpoints, and invariance to the manifestation of individual object features (Biederman 1987). To investigate whether this was the case, in a second “viewpoint contrast,” we measured whether the representation for the conjunctions changed when they were presented from a different viewpoint (i.e., were the conjunctive representations view-invariant?) (Fig. 2C, D). Importantly, in both contrasts, our novel MVPA linearity design avoided making unbalanced comparisons (e.g., $A + B + C$ vs. ABC) where the number of object features was confounded with the number of objects.

Indeed, an aspect of our design that should be emphasized is that during the task, participants viewed objects displayed in isolation. This is important, because presenting 2 objects simultaneously (e.g., Macevoy and Epstein 2009) could potentially introduce a bias, particularly when attention is divided between them (Reddy et al. 2009; Agam et al. 2010). Whereas objects were presented in isolation during the task, responses to single objects were then combined during analysis. On each side of every comparison, we combined across an equal number of objects (2), as there will be activity evoked by an object that does not scale with its number of features. So, for example, we rejected a simpler design in which $A + B = AB$ was tested, as there are an unequal number of objects combined on the 2 sides of the comparison (2 vs. 1).

We hypothesized that any observation of explicit conjunctive coding would be found in anterior VVS, extending into anterior temporal regions. In particular, one candidate structure that has received intensified interest is the perirhinal cortex (PRC)—a medial temporal lobe (MTL) structure whose function is traditionally considered exclusive to long-term memory (Squire and Zola-Morgan 1991; Murray and Bussey 2002; Bussey and Saksida 2002; Bussey and Saksida 2002; Bussey and Saksida 2002). Yet to our knowledge, there have been no direct investigations of explicit conjunctive coding in the PRC. Instead, most empirical attention has focused on area TE in monkeys and the object-selective

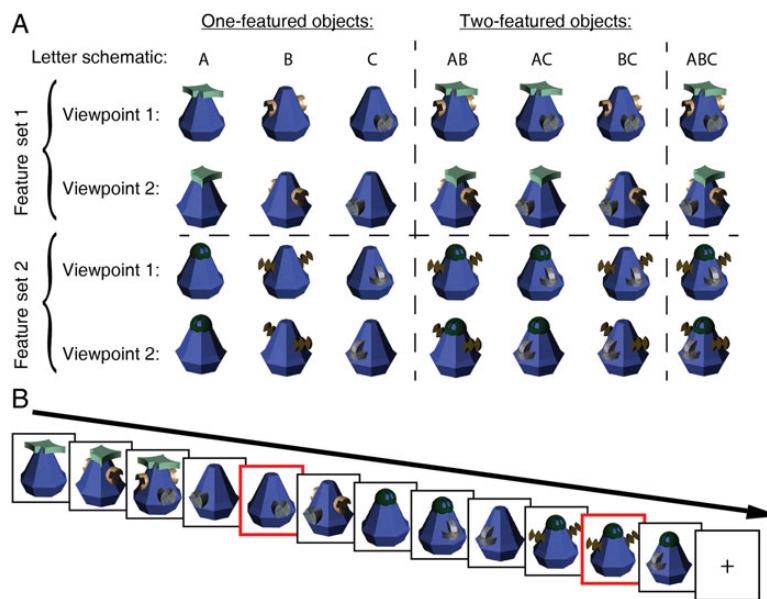


Figure 1. Stimuli and task. (A) Objects had 1, 2, or 3 attached features, and were shown from one of two possible viewpoints. For illustrative purposes, we schematize objects with letters: “A” corresponds to a one-featured object and “AB” corresponds to a two-featured object consisting of features A and B. We included 2 feature sets to ensure that our analysis did not yield results idiosyncratic to a particular feature set. (B) Participants completed a 1-back task in which they responded to a sequentially repeating object, regardless of viewpoint (targets in red). Objects were always presented in isolation.

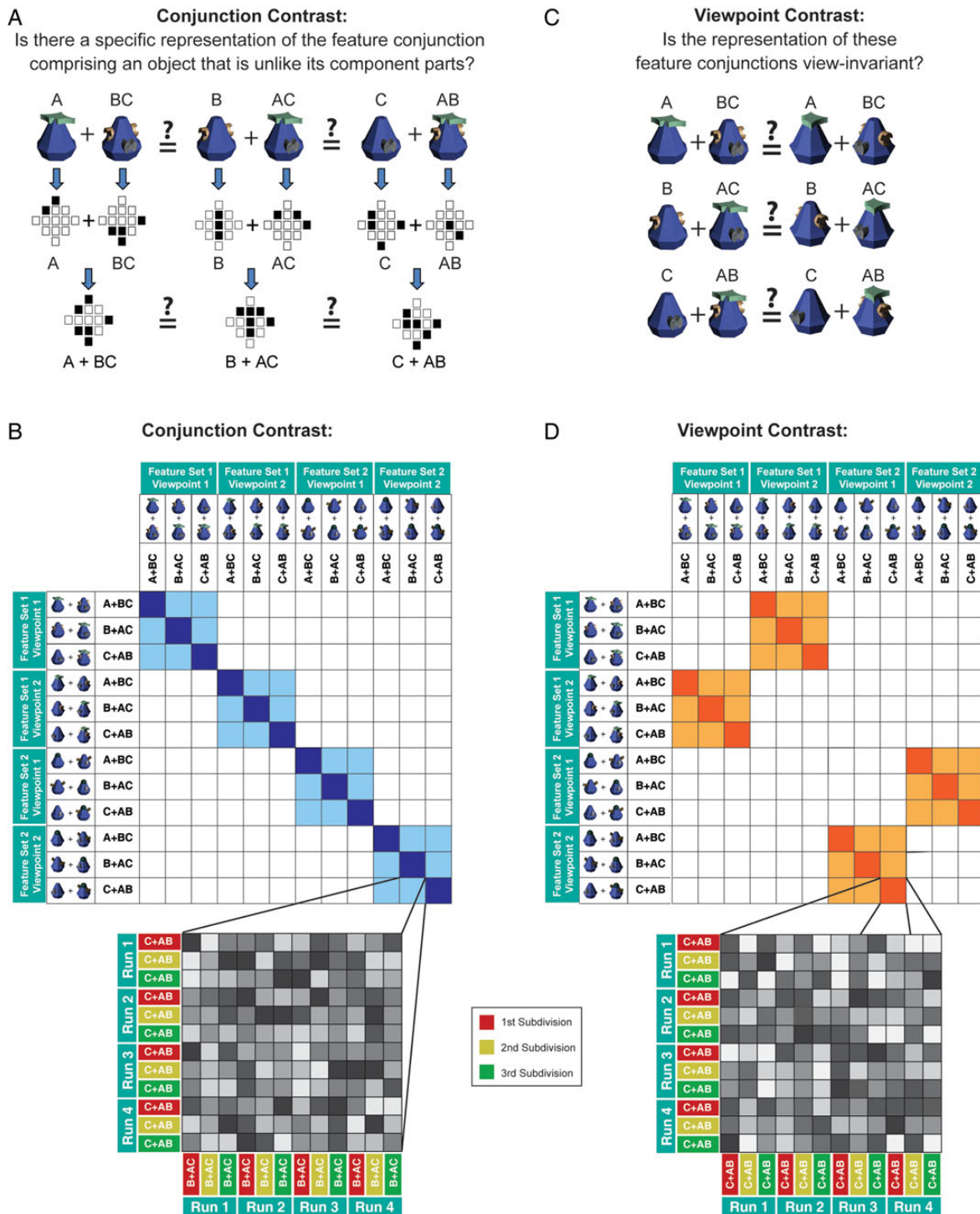


Figure 2. Experimental questions and similarity contrast matrices. (A) Our first contrast investigated whether neural patterns of activity demonstrated explicit conjunctive coding (i.e., was the whole different from the sum of its parts?). To this end, we measured patterns of activity (schematized by a hypothetical region of interest consisting of 13 voxels) to each of the objects that were presented individually during the 1-back task. We then computed linear summations of these patterns for 3 different pairs of objects (i.e., $A + BC$, $B + AC$, and $C + AB$), which were matched terms of their individual features ($A-C$), but different in terms of their conjunction (i.e., AB , BC , AC). (B) MVPA correlations within and between these conjunctions were summarized in a matrix structure (the full 144×144 matrix is shown in [Supplementary Fig. 2](#)). This contrast tested whether correlations of repetitions of the same conjunctions (dark blue) were more similar in their activation pattern compared with correlations of different conjunctions (light blue). As shown by the zoomed-in cell, each cell in the conjunction contrast is in fact an average of a 12×12 correlation matrix that computed correlations across the 4 experimental runs and the 3 data subdivisions. (C) Our second contrast investigated whether the conjunctive representations were view-invariant. As with the conjunction contrast (A), we measured the patterns of activity evoked by individually presented objects and computed linear summations of these patterns for each object pair. We then tested whether these patterns of activity were sensitive to the viewpoint at which the conjunction was presented. (D) MVPA correlations within and across the same conjunctions shown from different viewpoints were summarized in a matrix structure. This contrast tested whether correlations of repetitions of the same conjunctions presented from different viewpoints (dark orange) were more similar in their activation pattern compared with correlations between different conjunctions presented from different viewpoints (light orange). Note that the gray cells in the zoomed-in depiction reflect hypothetical data for illustrative purposes only.

lateral occipital complex (LOC) in humans—structures posterior to PRC and traditionally thought to be the anterior pinnacle of the VVS (Ungerleider and Haxby 1994; Grill-Spector et al. 2001; Denys et al. 2004; Sawamura et al. 2005; Kriegeskorte et al. 2008). For example, single-cell recording in monkey area TE showed evidence for conjunctive processing whereby responses to the whole object could not be predicted from the sum of the parts (Desimone et al. 1984; Baker et al. 2002; Gross 2008), although these conjunctive responses might have arisen from sensitivity to new features created by abutting features (Sripati and Olson 2010). Here, with our novel experimental design, we were able to directly measure explicit conjunctive coding of complex objects for the first time in humans. We used both a whole-brain approach and a region of interest (ROI)-based analysis that focused specifically on the PRC and functionally defined anterior structures in the VVS continuum. Our results revealed that regions of the VVS, extending into the PRC, contained unique representations of bound object features, consistent with an explicit conjunctive coding mechanism predicted by hierarchical models of object recognition.

Materials and Methods

Participants

Twenty neurologically normal right-handed participants gave written informed consent approved by the Baycrest Hospital Research Ethics Board and were paid \$50 for their participation. Data from one participant were excluded due to excessive head movement (>10° rotation), leaving 19 participants (18–26 years old, mean = 23.6 years, 12 females).

Experimental Stimuli and Design

Participants viewed novel 3D objects created, using Strata Design 3D CX 6. Each object was assembled from one of two feature sets and was composed of a main body with 1, 2, or 3 attached features (depicted as “A,” “B,” and “C” in Fig. 1A). There were 7 possible combinations of features within a feature set (A, B, C, AB, AC, BC, ABC). Features were not mixed between sets. Each object was presented from one of two possible angles separated by a 70° rotation along a single axis: 25° to the right, and 45° to the left from central fixation. We ensured that all features were always visible between angle changes. There were 28 images in total, created from every unique combination of the experimental factors: 2 (feature sets) × 2 (viewpoints) × 7 (possible combinations of features within a set). Figure 1A depicts the complete stimulus set.

Image Similarity Analysis

To measure the basic visual similarity between the objects in our stimulus set, we calculated the root-mean-square difference (RMSD) between each of the 24 objects (all one-featured and two-featured objects) and compared this value to every other object using the following function:

$$\sqrt{\frac{\sum_{i=1}^n (\text{Image}1_i - \text{Image}2_i)^2}{n}}$$

where i is a pixel position in the image and n is the total number of pixels in the image. Thus, this function compares all of the pixels in 1 image with the corresponding pixels in a second image and yields a value that indicates the similarity between 2 images, ranging from 0 (if the 2 images are identical) to 1 (if the 2 images are completely different) (Fig. 3A). The purpose of this analysis

was to determine how similar 2 images are on the most basic of levels—that is, how different the images would appear to the retina. Specifically, we conducted an analysis to ensure that our viewpoint manipulation in fact caused a substantial change in the visual appearance of the objects. For example, if our viewpoint shifts were insignificant (e.g., 1° difference), any observation of view-invariance for this small visual change would not be particularly meaningful. However, if we could show that our shift in viewpoint caused a visual change that was as significant as a change in the identity of the object itself, a demonstration of view-invariance would be much more compelling. To this end, we calculated a contrast matrix (Fig. 3B) that compared the RMSD values of the same objects from different viewpoints (dark orange) versus RMSD values of different object features from the same viewpoint (light orange). Different objects from the same viewpoint were compared only if they were matched for the number of features. A t -test revealed that there was a significant difference between RMSD values of the same features shown from different viewpoints ($M = 0.16$, $SD = 0.03$) compared with different features shown from the same viewpoint ($M = 0.11$, $SD = 0.02$; $t_{(34)} = 6.11$, $P < 0.001$), revealing that that our viewpoint manipulation caused a change in visual appearance that was more drastic than maintaining the same viewpoint but changing the identity of the features altogether. Next, we conducted an RMSD analysis that was very similar to the viewpoint contrast (shown in Fig. 2D) used in our MVPA. Here, we compared the RMSD values of a change in viewpoint (dark orange) with a change in both viewpoint and feature type (light orange) (Fig. 3C). A t -test revealed that there was no significant difference between RMSD values of the same features shown from different viewpoints ($M = 0.16$, $SD = 0.03$) compared with different features shown from different viewpoints ($M = 0.17$, $SD = 0.02$; $t_{(70)} = 1.28$, $P = 0.20$). This indicates that the shift in viewpoint within a feature set caused a change in visual appearance that was as drastic as changing the features altogether. That is, our viewpoint manipulation was not trivial and caused a substantial visual change of the appearance of the objects.

It is worth noting that because RMSD values constitute a difference score that reflects the visual differences between one object with respect to another object, we could not calculate RMSD difference scores to the pairs of objects as we did in the MVPA (e.g., A + BC vs. B + AC). Put differently, in the MVPA we could measure the patterns of activity evoked by the presentation of a single object (e.g., A), add this pattern of activity to that evoked by a different single object (e.g., BC), and then compare across different pattern sums (e.g., A + BC vs. B + AC). In contrast, because an RMSD value reflects the visual difference between 2 objects (rather than to an object on its own), we could not measure an RMSD value to object “A” and add that value to the RMSD value of object “BC.” As such, it was not possible to calculate RMSD values for the pattern sums.

Tasks

Experimental Task

We administered 4 experimental scanning runs during which participants completed a 1-back task to encourage attention to each image. Participants were instructed to press a button with their right index finger whenever the same object appeared twice in succession, regardless of its viewpoint (behavioral results in Supplementary Table 1). Feedback was presented following each button press (correct or incorrect) and at the end of each run (proportion of correct responses during that run). Trials on which a response was made were not included in the analysis.

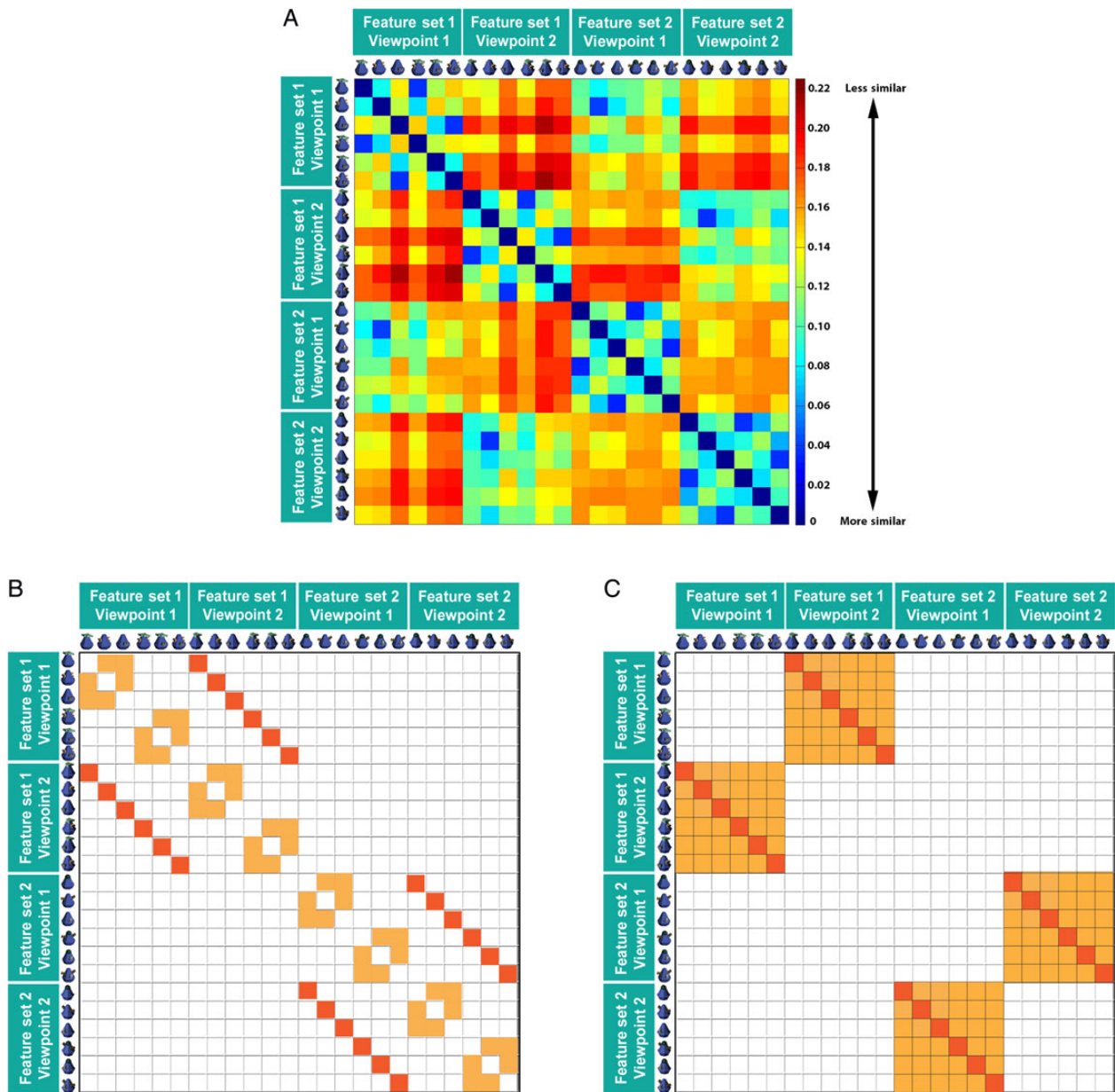


Figure 3. An analysis of visual differences across objects using a root-mean-square difference (RMSD) measure on the RGB values of all images used in the experiment indicated that the viewpoint manipulation produced a substantial change in the visual appearance of the object that was as drastic as switching the identity of the object altogether. (A) Raw RMSD values for all images used in the experiment. RMSD values range from 0 (if the 2 images were identical) to 1 (if the 2 images were completely different, see bar on the right). (B) A contrast matrix that compared the RMSD values of the same objects from different viewpoints (dark orange) versus RMSD values of different objects from the same viewpoint (light orange). Objects from the same viewpoint were compared only if they were matched for the number of features. (C) Similar to the viewpoint contrast (Fig. 2D) used in our MVPA, we constructed a contrast matrix that compared the RMSD values of a change in viewpoint (dark orange) to a change in both viewpoint and feature type (light orange).

Objects were presented centrally on the screen and had a visual angle of $5.1^\circ \times 5.3^\circ$, which would likely encompass the receptive fields of PRC ($\sim 12^\circ$) (Nakamura et al. 1994), V4 ($4\text{--}6^\circ$ at an eccentricity of 5.5°) (Kastner et al. 2001), LOC ($4\text{--}8^\circ$) (Dumoulin and Wandell 2008), and fusiform face area (FFA), and parahippocampal place area (PPA) (likely $>6^\circ$) (Desimone and Duncan 1995; Kastner et al. 2001; Kornblith et al. 2013). The visual angle of the individual object features was approximately $2.1^\circ \times 2.2^\circ$, which would likely encompass the receptive fields of more posterior regions in the VVS ($2\text{--}4^\circ$ in V2) (Kastner et al. 2001). Each image was displayed for 1 s with a 2 s interstimulus interval. Each run

lasted 11 min 30 s, and for every 42 s of task time, there was an 8 s break (to allow blood oxygen level dependent (BOLD) signal to reach baseline) during which a fixation cross appeared on the screen. Each run comprised 6 blocks of 28 trials, which were presented in a different order to each participant. The 14 images composing each feature set were randomly presented twice within each block. Across consecutive blocks, the feature sets alternated (3 blocks per feature set per run). Each block contained between 1 and 4 target objects (i.e., sequential repeats), such that the overall chance that an object was a target was 10%. In total, each image was presented 24 times (6 times within

each run). Prior to scanning, each participant performed a 5-min practice of 60 trials.

Localizer Task

After the 4 experimental runs, an independent functional localizer was administered to define participant-specific ROIs (LOC, FFA, and PPA, described next). Participants viewed scenes, faces, objects, and scrambled objects in separate 15-s blocks (there was no overlap between the images in the experimental task above and the localizer task). Within each block, 20 images were presented for 300 ms each with a 450-ms ISI. There were 4 groups of 12 blocks, with each group separated by a 15-s fixation-only block. Within each group, 3 scene, face, object, and scrambled object blocks were presented (order of block type was counterbalanced across groups). To encourage attention to each image, participants were instructed to press a button with their right index finger whenever the same image appeared twice in succession. Presentation of images within blocks was pseudo-random: immediate repeats occurred between 0 and 2 times per block.

Memory Task

Following scanning, participants were administered a memory task in which they determined whether a series of objects were shown during scanning (Supplementary Fig. 1 for description of the task and results). Half of the objects were seen previously in the scanner and half were novel recombinations of features from across the 2 feature sets. In brief, the results indicated that participants could discriminate easily between previously viewed objects and objects comprising novel reconfigurations of features, suggesting that the binding of features extended beyond the immediate task demands in the scanner, but also transferred into longer-term memory.

fMRI Data Acquisition

Scanning was performed using a 3.0-T Siemens MAGNETOM Trio MRI scanner at the Rotman Research Institute at Baycrest Hospital using a 32-channel receiver head coil. Each scanning session began with the acquisition of a whole-brain high-resolution magnetization-prepared rapid gradient-echo T_1 -weighted structural image (repetition time = 2 s, echo time = 2.63 ms, flip angle = 9°, field of view = 25.6 cm², 160 oblique axial slices, 192 × 256 matrix, slice thickness = 1 mm). During each of four functional scanning runs, a total of 389 T_2^* -weighted echo-planar images were acquired using a two-shot gradient echo sequence (200 × 200 mm field of view with a 64 × 64 matrix size), resulting in an in-plane resolution of 3.1 × 3.1 mm for each of 40 2-mm axial slices that were acquired along the axis of the hippocampus. The interslice gap was 0.5 mm; repetition time = 2 s; echo time = 30ms; flip angle = 78°.

Multivoxel Pattern Analysis

Functional images were preprocessed and analyzed using SPM8 (www.fil.ion.ucl.ac.uk/spm) and a custom-made, modular toolbox implemented in an automatic analysis pipeline system (<https://github.com/rhodricusack/automaticanalysis/wiki>). Prior to MVPA, the data were preprocessed, which included realignment of the data to the first functional scan of each run (after 5 dummy scans were discarded to allow for signal equilibrium), slice-timing correction, coregistration of functional and structural images, nonlinear normalization to the Montreal Neurological Institute (MNI) template brain, and segmentation of gray and white matter. Data were high-pass filtered with a 128-s

cutoff. The data were then “denoised” by deriving regressors from voxels unrelated to the experimental paradigm and entering these regressors in a general linear model (GLM) analysis of the data, using the GLM denoise toolbox for Matlab (Kay et al. 2013). Briefly, this procedure includes taking as input a design matrix (specified by the onsets for each stimulus regardless of its condition) and an fMRI time-series, and returns as output an estimate of the hemodynamic response function (HRF) and BOLD response amplitudes (β weights). It is important to emphasize that the design matrix did not include the experimental conditions upon which our contrasts relied; these conditions were specified only after denoising the data. Next, a fitting procedure selected voxels that are unrelated to the experiment (cross-validated $R^2 < 0\%$), and a principal components analysis was performed on the time-series of these voxels to derive noise regressors. A cross-validation procedure then determined the number of regressors that were entered into the model (Kay et al. 2013).

We specified the onsets for each individual object (i.e., A, B, C, AB, BC, AC) for each of the 2 feature sets and 2 viewpoints. Our model then created a single regressor for each of the three different pairs of objects (i.e., A + BC, B + AC, and C + AB). This was done separately for each of the 2 feature sets and 2 viewpoints. For example, events corresponding to the singly presented “A” object from feature set 1 and viewpoint 2 and events corresponding to the singly presented “BC” object from feature set 1 and viewpoint 2 were combined to create the single regressor for “A + BC” from feature set 1 and viewpoint 2. More specifically, within each run, the voxel-wise data of each object pair were split into 3 subdivisions that were each composed of every third trial of a given image (following Zeithamova et al. 2012) (Fig. 2B,D, zoomed-in cells). The pattern similarity of each condition in each subdivision was compared with that of each condition in every other subdivision. We designed the subdivisions so that our comparisons were relatively equidistant in time. For example, the first subdivision for the A + BC regressor included $A_{1st\ presentation} + BC_{1st\ presentation} + A_{4th\ presentation} + BC_{4th\ presentation}$; the second subdivision included $A_{2nd\ presentation} + BC_{2nd\ presentation} + A_{5th\ presentation} + BC_{5th\ presentation}$, etc. This resulted in 36 regressors of interest per run [2 (feature sets) × 2 (viewpoints) × 3 (conjunctions) × 3 (subdivisions)]. We also modeled 8 regressors of no interest for each run: trials of three-featured objects (ABC), trials in which participants responded with a button press on the 1-back task, and 6 realignment parameters to correct for motion. Events were modeled with a delta (stick) function corresponding to the stimulus presentation onset convolved with the canonical HRF as defined by SPM8. This resulted in parameter estimates (β) indexing the magnitude of response for each regressor. Multivoxel patterns associated with each regressor were then Pearson-correlated. Thus, each cell in our planned contrast matrices was composed of a 12 × 12 correlation matrix that computed correlations within and across all runs and data subdivisions (Fig. 2B,D, zoomed-in cells; see also Supplementary Fig. 2 for the full data matrix). This process was repeated for each cell in the contrast matrix, and these correlation values were then averaged and condensed to yield the 12 × 12 contrast matrix (similar to Linke et al. 2011). We then subjected these condensed correlation matrices to our planned contrasts (Fig. 2B,D).

In addition to an analysis that computed correlations both across and within runs, we also conducted an analysis in which we ignored within-run correlations and computed correlations across runs only. Results from this across-run only analysis are shown in Supplementary Figure 3. In brief, this analysis revealed the same pattern of results as the analysis that computed correlations both across and within runs.

Searchlight Analysis

A spherical ROI (10 mm radius, restricted to gray matter voxels and including at least 30 voxels) was moved across the entire acquisition volume, and for each ROI, voxel-wise, unsmoothed β -values were extracted separately for each regressor (Krieges-korte et al. 2006). The voxel-wise data (i.e., regressors of interest) were then Pearson-correlated within and across runs, and condensed into a 12×12 correlation matrix (see Fig. 2B,D). Predefined similarity contrasts containing our predictions regarding the relative magnitude of pattern correlations within and between conjunction types specified which matrix elements were then subjected to a two-sample t-test. This analysis was performed on a single-subject level, and a group statistic was then calculated from the average results, indicating whether the ROI under investigation coded information according to the similarity matrix. Information maps were created for each subject by mapping the t-statistic back to the central voxel of each corresponding ROI in that participant's native space. These single-subject t-maps were then normalized, and smoothed with a 12-mm full width at half maximum (FWHM) Gaussian kernel to compensate for anatomical variability across participants. The resulting contrast images were then subjected to a group analysis that compared the mean parameter-estimate difference across participants to zero (i.e., a one-sample t-test relative to zero). Results shown in Figure 4A,C are superimposed on the single-subject MNI brain template.

ROI Analysis

We investigated 4 ROIs defined a priori. Three were functionally defined regions well established as part of the VVS: LOC, FFA, and the PPA. The fourth ROI was the PRC, which was defined by an anatomical probability map created by Devlin and Price (2007). We included areas, which had at least a 30% or more probability of being the PRC, as done previously (Barense et al. 2011). For our functional localizer, we used identical stimuli to those employed in Watson et al. (2012). We defined the LOC as the region that was located along the lateral extent of the occipital lobe and responded more strongly to objects compared with scrambled objects ($P < 0.001$, uncorrected) (Malach et al. 1995). We defined the FFA as the set of contiguous voxels in the mid-fusiform gyrus that showed significantly higher responses to faces compared with objects ($P < 0.001$, uncorrected) (Liu et al. 2010), and the PPA as the set of contiguous voxels in the parahippocampal gyrus that responded significantly more to scenes than to objects ($P < 0.001$, uncorrected) (Reddy and Kanwisher 2007). These regions were defined separately for each participant by a 10-mm radius sphere centered around the peak voxel in each hemisphere from each contrast, using the MarsBar toolbox for SPM8 (<http://marsbar.sourceforge.net/>). All ROIs were bilateral, except for 3 participants in whom the left FFA could not be localized, and another participant in whom the right LOC could not be localized (Supplementary Table 2 displays the peak ROI coordinates for each ROI for each participant). The ROI MVPA was conducted in an identical manner to the searchlight analysis; voxel-wise data were Pearson-correlated and condensed into a 12×12 correlation matrix, except that here each ROI was treated as a single region (i.e., no searchlights were moved within an ROI). Before applying our contrasts of interest, we ensured that these correlation values were normally distributed (Jarque-Bera test; $P > 0.05$). We then applied our conjunction and viewpoint contrasts within each of the four ROIs and obtained, for each participant and each contrast, a t-value reflecting the strength of the difference between our correlations of interest (Fig. 2B,D). From

these t-values, we calculated standard r -effect sizes that allowed us to compare the magnitude of effects across the ROIs (Rosenthal 1994) (Fig. 4B,D). Specifically, we transformed the r -effect sizes to Fisher's z-scores (as they have better distribution characteristics than correlations, e.g., Mullen 1989). We then conducted t-tests on the z-scores effect sizes obtained for each region, which provided a measure of the statistical significance between our cells of interest in each of our two contrast matrices (i.e., dark- and light-colored cells). We then compared the z-scores in each ROI to zero using Bonferroni-corrected one-sample t-tests, and conducted for each of the two contrasts paired-samples t-tests to compare the effect sizes observed in the PRC to the 3 more posterior ROIs (Fig. 4B,D).

Control Analysis to Test for Subadditivity of the BOLD Signal

To test the possibility that our results (Fig. 4) were driven by signal saturation due to nonlinearities in neurovascular or vascular MR coupling, we conducted a univariate ANOVA to test whether there were differences in the overall signal evoked by different single features (Fig. 5). A standard univariate processing pipeline was then followed, comprising the preprocessing steps described for the MVPA, but also smoothing of the imaging data with a 12-mm FWHM Gaussian kernel. We then conducted first-level statistical analyses. Within each run, there were 12 regressors of interest [2 (feature sets) $\times 2$ (viewpoints) $\times 3$ objects (A, B, and C)] and 8 regressors of no interest corresponding to trials of three-featured objects (ABC), trials in which participants responded with a button press on the 1-back task, and 6 realignment parameters to correct for motion. Within each regressor, events were modeled by convolving a delta (stick) function corresponding to the stimulus presentation onset with the canonical HRF as defined by SPM8. Second-level group analyses were conducted separately for each of the two feature sets by entering the parameter estimates for the 6 single-featured objects (i.e., A, B, and C from each viewpoint) of each subject into a single GLM, which treated participants as a random effect. This analysis was conducted using a factorial repeated-measures ANOVA, in which a model was constructed for the main effect of condition (i.e., the 6 single features). Within this model, an F-contrast was computed to test for areas that showed a significant effect of feature type for each feature set. Statistical parametric maps (SPMs) of the resulting F-statistic were thresholded at $P < 0.001$, uncorrected.

Results and Discussion

Our primary analyses involved 2 planned comparisons. The first comparison, the conjunction contrast, determined whether the neural patterns of activity demonstrated explicit conjunctive coding (i.e., whether activity patterns represented information specific to the conjunction of features comprising an object, over and above information regarding the features themselves) (Fig. 2A,B). The second comparison, the viewpoint contrast, investigated whether the conjunctive representations were view-invariant (Fig. 2C,D). For each of these two contrasts, we performed 2 independent planned analyses—a “searchlight analysis” to investigate the activity across the whole brain and an “ROI analysis” to investigate activation in specific VVS ROIs. Both contrasts were applied to a correlation matrix that included all possible correlations within and across the different conjunctions (Fig. 2B,D). The novelty of this design ensured that our comparisons were matched in terms of the number of features that

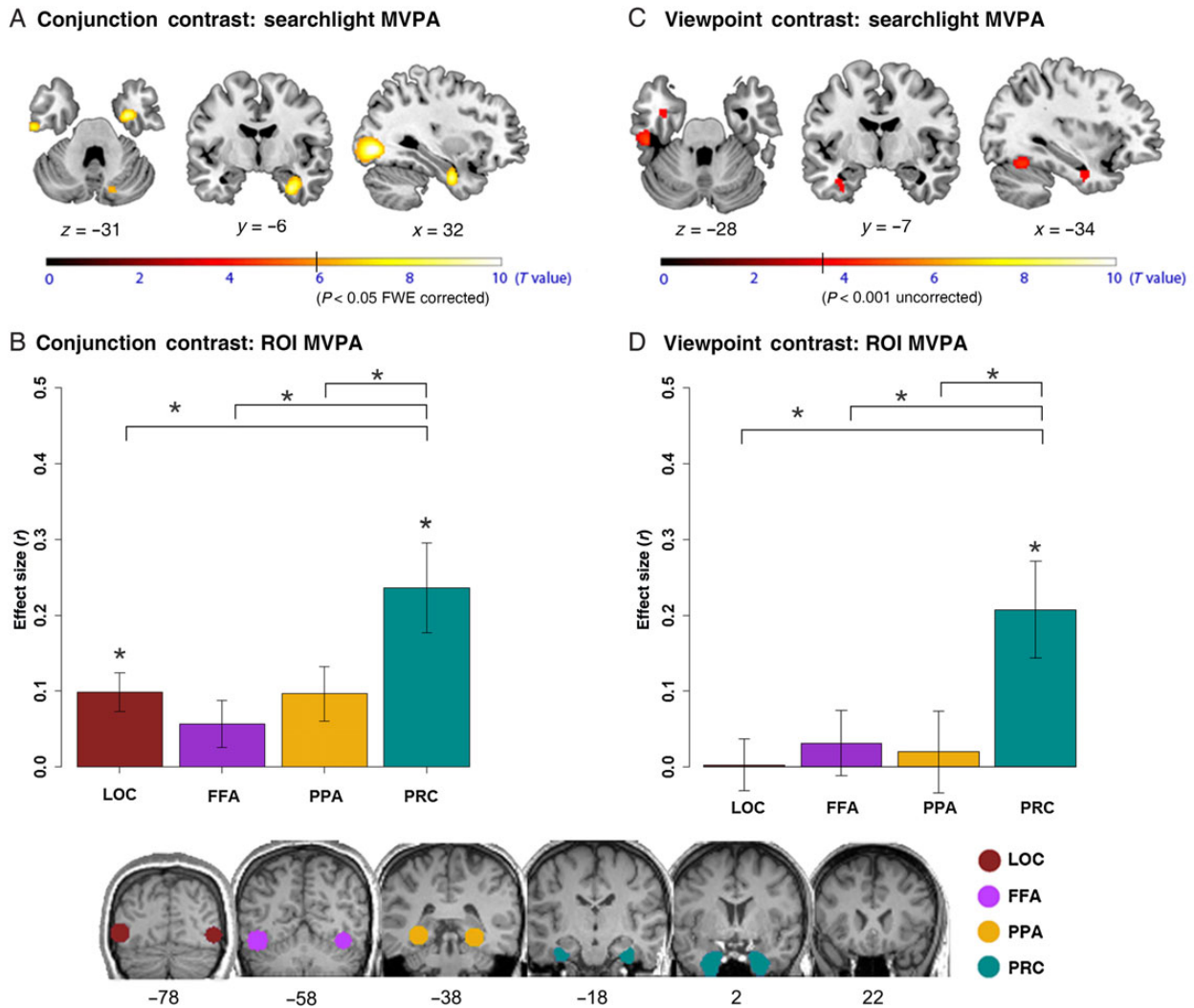


Figure 4. Regions demonstrating explicit conjunctive coding within and across viewpoints. (A) Regions where the representation for the conjunction was different from the sum of its parts ($P < 0.05$, whole-brain FWE-corrected; results at $P < 0.001$, uncorrected shown in [Supplementary Fig. 5](#)). Broadly speaking, these regions included the PRC and earlier VVS regions (with the local maxima in V4). (B) The strength of conjunctive coding in the LOC, FFA, PPA, and PRC ROIs (shown at bottom are ROIs from a representative participant superimposed on that individual's structural scan). (C) Regions demonstrating view-invariant conjunctive representations ($P < 0.001$, uncorrected; no VVS regions survived whole-brain FWE correction). Broadly speaking, these regions included the PRC, as well as V4 and lateral IT cortex. (D) The strength of view-invariant conjunctive coding within and across ROIs. Error bars indicate SEM; * $P < 0.05$ for comparisons across ROIs (paired-samples *t*-tests indicated by brackets) and comparisons relative to zero within an ROI (one-sample *t*-tests indicated by an asterisk above a bar).

needed to be bound. That is, our comparison terms (e.g., A + BC vs B + AC) included both a combination of a single-featured object and a two-featured object, and thus, binding and memory requirements were matched—what differed was the underlying representation for the conjunctions themselves.

For the conjunction contrast, a whole-brain searchlight MVPA ([Kriegeskorte et al. 2006](#)) revealed conjunctive coding in the VVS, with a global maxima in V4 ([Rottschy et al. 2007](#)) and activity that extended laterally into the LOC and posteriorly into V3 and V1 (peak $x, y, z = 32, -80, -4$, Z -value = 5.93), as well as conjunctive coding that extended anteriorly to the right PRC (Fig. 4A; peak $x, y, z = 30, -6, -30$, Z -value = 5.47) (all results reported are whole-brain FWE-corrected at $P < 0.05$; [Supplementary Table 3](#) summarizes all regions). We next performed an ROI-based MVPA that applied the same contrast matrices and methods used for the searchlight analysis, except was focused only on the PRC and 3 functionally defined regions (LOC, FFA, and PPA) that are posterior to the PRC and are well established as part of the VVS.

This analysis allowed direct comparison of conjunctive coding strength across regions (Fig. 4B). The conjunction contrast in this ROI MVPA revealed conjunctive coding in PRC ($t_{(18)} = 3.89$, $P < 0.01$, $r_{\text{effect size}} = 0.24$) and the LOC ($t_{(18)} = 3.85$, $P < 0.01$, $r_{\text{effect size}} = 0.10$), but not in FFA ($t_{(18)} = 1.80$, $P = 0.35$, $r_{\text{effect size}} = 0.06$), or PPA ($t_{(18)} = 2.66$, $P = 0.06$, $r_{\text{effect size}} = 0.10$) (all one-sample *t*-tests Bonferroni-corrected). Comparisons across ROIs demonstrated stronger conjunctive coding in the PRC relative to each of the three more posterior VVS ROIs (P 's < 0.05). Thus, consistent with recent hierarchical models proposing explicit conjunctive coding in regions not traditionally associated with the VVS ([Murray et al. 2007](#); [Barense, Groen, et al. 2012](#)), we found that PRC representations explicitly coded information regarding the object's conjunction, over and above its individual features.

In addition to the PRC, we also observed conjunctive coding in V4 as well as in LOC, indicating that the conjunctive coding mechanism is not selective to PRC. Indeed, there is evidence to suggest that V4 and LOC are important for representing feature

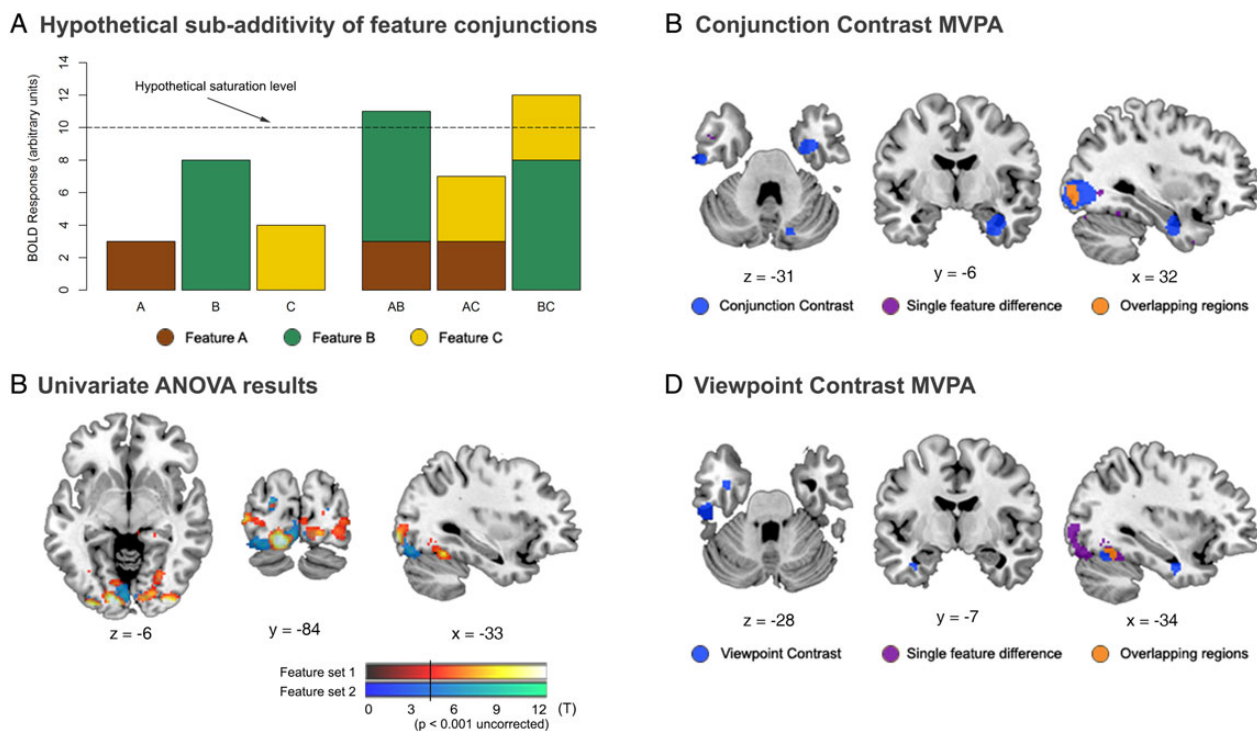


Figure 5. Control analysis to test for subadditivity of the blood oxygen level dependent (BOLD) signal. (A) Saturation of the BOLD signal could produce nonlinearities that could be misinterpreted as a conjunction—when in fact the coding was linear—if there were stronger activation for some features more than others (e.g., illustrated in the earlier hypothetical scenario for feature B). In this instance, adding another feature (e.g., A, C) to the feature that already produces strong activity (B) could cause the BOLD signal to saturate (AB, BC: dotted line). The BOLD response to the feature pairs would be a nonlinear addition of that to the component features—not because of a conjunctive neural representation, but rather due to sub-additivity of the BOLD signal. (B) Regions identified by our univariate ANOVA as showing a difference in terms of overall signal evoked by the different single features in either feature set 1 (warm colors) or feature set 2 (cool colors). This analysis revealed activity in posterior visual regions, but not in PRC (see [Supplementary Table 5](#)). Searchlight MVPA results from the (C) conjunction contrast (all regions from Fig. 4A shown here in blue) and (D) viewpoint contrast (all regions from Fig. 4C shown in blue), displayed with regions from the combined results of the univariate ANOVA for feature sets 1 and 2 (shown in purple). Overlapping regions between the 2 analyses are shown in orange.

conjunctions. For example, recent studies showed that learning-induced performance changes on a conjunctive visual search were correlated with increasing activity in V4 and LOC (Frank et al. 2014), and that these regions are important for conjoined processing of color and spatial frequency (Pollmann et al. 2014). In support of a causal role for both LOC and PRC in representing feature conjunctions, patients with selective LOC damage (Behrmann and Williams 2007; Konen et al. 2011) and those with PRC damage (Barens et al. 2005; Barens, Groen, et al. 2012) were impaired on tasks that tax integrating object features into a cohesive unit. Nonetheless, we did observe clear differences in conjunctive versus single-feature coding across our ROIs, with PRC demonstrating stronger conjunctive coding than LOC, FFA, and PPA (Fig. 4B), and LOC demonstrating stronger single-feature coding than PRC, FFA, or PPA ([Supplementary Fig. 4](#)). Although the current study lacks the temporal resolution to address this directly, one possibility is that this more posterior activity may also reflect higher level feedback, such as from PRC. Indeed, bidirectional interactions exist throughout the VVS (Hochstein and Ahissar 2002; Coutanche and Thompson-Schill 2015), and previous work has suggested that feedback from the PRC modulates familiarity responses to object parts in V2 (Barens, Ngo, et al. 2012; Peterson et al. 2012). An alternative possibility is that the PRC activity reflects a feedforward cascade from structures such as the LOC (Lamme and Roelfsema 2000; Bullier 2001).

Next, we investigated whether the conjunctive representations of the object features were view-invariant. The extent to which the human visual system supports view-invariant versus

view-dependent representations of objects is unresolved (Biederman and Bar 1999; Peissig and Tarr 2007). Much research in this area has focused on VVS regions posterior to the MTL (Vuilleumier et al. 2002; Andresen et al. 2009) but some work has indicated that, for very complex stimuli, structures in the MTL may be central to view-invariance (Quiroga et al. 2005; Barens et al. 2010). Despite this, to our knowledge, no study has directly probed how the specific conjunctions comprising complex objects are neurally represented across viewpoints to support object recognition in different viewing conditions. That is, as the representations for objects become increasingly precise and dependent on the specific combinations of features comprising them, can they also become more invariant to the large changes in visual appearance caused by shifting viewpoints?

To investigate this question, we presented the objects at 70° rotations (Fig. 1), a manipulation that caused a more drastic visual change than changing the identity of the object itself (Fig. 3). To assess view-invariance, our viewpoint contrast compared the additivity of patterns (i.e., A + BC vs. B + AC vs. C + AB) across the 2 viewpoints (Fig. 2C,D). At a stringent whole-brain FWE-corrected threshold of $P < 0.05$, our searchlight MVPA revealed limited activity throughout the brain (2 voxels, likely in the orbitofrontal cortex, [Supplementary Table 4A](#)). However, at a more liberal threshold ($P < 0.001$ uncorrected), we observed view-invariant conjunctive coding in the VVS, with maxima in V4 (Rottschy et al. 2007); peak $x,y,z = 32, -70, -4$, Z -value = 4.07), lateral IT cortex (peak $x,y,z = -52, -28, -20$, Z -value = 4.15), as well as activity that extended anteriorly to the left PRC (peak $x,y,z = -36, -4, -26$,

Z-value = 3.25) (Fig. 4C and Supplementary Table 4B). The ROI MVPA of the viewpoint contrast revealed view-invariance in the PRC ($t_{(18)} = 3.04$, $P < 0.05$, $r_{\text{effect size}} = 0.21$), but not the LOC, FFA, or PPA (t 's < 0.75 , P 's > 0.99 , all $r_{\text{effect size}} < 0.03$; Fig. 4D) (all tests Bonferroni-corrected). A direct comparison across regions confirmed that this view-invariant conjunctive coding was stronger in the PRC compared with the 3 more posterior VVS ROIs (P 's < 0.05).

Taken together, results from the conjunction and viewpoint contrasts provide the first direct evidence that information coding in PRC simultaneously discriminated between the precise conjunctions of features comprising an object, yet was also invariant to the large changes in the visual appearance of those conjunctions caused by shifting viewpoints. These coding principles distinguished the PRC from other functionally defined regions in the VVS that are more classically associated with perception (e.g., LOC, FFA, and PPA). It is important to note that our results were obtained in the context of a 1-back memory task, and thus, one might be cautious to interpret our results in the context of a perceptual role for the PRC. Indeed, we prefer to consider our results in terms of the *representational* role for any given brain region, rather than in terms of the cognitive process it supports—be it either memory or perception. To this end, our experimental design ensured that memory demands were matched in our comparisons of interest—all that differed was how the same 3 features were arranged to create an object. With this design, explicit memory demands were equated, allowing a clean assessment of the underlying object representation. That said, one might still argue that these object representations were called upon in the service of a 1-back memory task. However, a wealth of evidence suggests that PRC damage impairs performance on both perceptual (e.g., Bussey et al. 2002; Bartko et al. 2007; Lee and Rudebeck 2010; Barense, Ngo, et al. 2012; Barense et al. 2012) and explicit memory tasks (Bussey et al. 2002; Winters et al. 2004; Barense et al. 2005; Bartko et al. 2010; McTighe et al. 2010). The critical factor in eliciting these impairments was whether the task required the objects to be processed in terms of their conjunctions of features, rather than on the basis of single features alone. We argue that these seemingly disparate mnemonic and perceptual deficits can be accounted for by the fact that both the mnemonic and perceptual tasks recruited the conjunctive-level representations we have measured in the current study.

Finally, it is important to rule out the possibility that the results we obtained were driven by signal saturation due to nonlinearities in neurovascular or vascular MR coupling (Fig. 5). Such signal saturation could produce nonlinearities that could be misinterpreted as a conjunction, when in fact the coding was linear, if there were stronger activation for some features more than others (Fig. 5A). To evaluate whether our data fell within this regime, we conducted a univariate ANOVA to test whether there were differences in the overall signal evoked by different features when these features were presented alone. These analyses revealed activity predominately in early visual areas that were largely nonoverlapping with the results from our critical conjunction and viewpoint contrasts (Fig. 5C,D and Supplementary Table 5). Importantly, there were no significant differences between the basic features in terms of overall activity in the PRC, or in the LOC typically observed in our participants (even at a liberal uncorrected threshold of $P < 0.001$). Thus, there was no evidence to suggest our observation of conjunctive coding is driven by spurious BOLD signal saturation.

In conclusion, this study provides new evidence regarding the functional architecture of object perception, demonstrating neural representations that both integrated object features into

a whole that was different from the sum of its parts and were also invariant to large changes in viewpoint. To our knowledge, this constitutes the first direct functional evidence for explicit coding of complex feature conjunctions in the human PRC, suggesting that visual-object processing does not culminate in IT cortex as long believed (e.g., Lafer-Sousa and Conway 2013), but instead continues into MTL regions traditionally associated with memory. This is consistent with recent proposals that memory (MTL) and perceptual (VVS) systems are more integrated than previously appreciated (Murray et al. 2007; Barense, Groen, et al. 2012). Rather than postulating anatomically separate systems for memory and perception, these brain regions may be best understood in terms of the representation they support—and any given representation may be useful for many different aspects of cognition. In the case of the PRC, damage to these complex conjunctive representations would not only impair object perception, but would also cause disturbances in the recognition of objects and people that are critical components of amnesia.

Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org>

Funding

This work was supported by the Canadian Institutes of Health Research (MOP-115148 to M.D.B.) and a Scholar Award from the James S. McDonnell Foundation to M.D.B.

Notes

We thank Christopher Honey and Andy Lee for their helpful comments on an earlier version of this manuscript, and Nick Rule and Alejandro Vicente-Grabovetsky for assistance with data analysis. *Conflict of Interest:* None declared.

References

- Agam Y, Liu H, Papanastassiou A, Buia C, Golby AJ, Madsen JR, Kreiman G. 2010. Robust selectivity to two-object images in human visual cortex. *Curr Biol*. 20:872–879.
- Andresen DR, Vinberg J, Grill-Spector K. 2009. The representation of object viewpoint in human visual cortex. *Neuroimage*. 45:522–536.
- Baker CI, Behrmann M, Olson CR. 2002. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nat Neurosci*. 5:1210–1216.
- Barense MD, Bussey TJ, Lee ACH, Rogers TT, Davies RR, Saksida LM, Murray EA, Graham KS. 2005. Functional specialization in the human medial temporal lobe. *J Neurosci*. 25:10239–10246.
- Barense MD, Groen II, Lee ACH, Yeung LK, Brady SM, Gregori M, Kapur N, Bussey TJ, Saksida LM, Henson RNA. 2012. Intact memory for irrelevant information impairs perception in amnesia. *Neuron*. 75:157–167.
- Barense MD, Henson RNA, Graham KS. 2011. Perception and conception: temporal lobe activity during complex discriminations of familiar and novel faces and objects. *J Cogn Neurosci*. 23:3052–3067.
- Barense MD, Henson RNA, Lee ACH, Graham KS. 2010. Medial temporal lobe activity during complex discrimination of faces, objects, and scenes: effects of viewpoint. *Hippocampus*. 20:389–401.

- Barense MD, Ngo JKW, Hung LHT, Peterson MA. 2012. Interactions of memory and perception in amnesia: the figure-ground perspective. *Cereb Cortex*. 22:2680–2691.
- Bartko SJ, Cowell RA, Winters BD, Bussey TJ, Saksida LM. 2010. Heightened susceptibility to interference in an animal model of amnesia: impairment in encoding, storage, retrieval—or all three? *Neuropsychologia*. 48:2987–2997.
- Bartko SJ, Winters BD, Cowell RA, Saksida LM, Bussey TJ. 2007. Perirhinal cortex resolves feature ambiguity in configural object recognition and perceptual oddity tasks. *Learn Mem*. 14:821–832.
- Behrmann M, Williams P. 2007. Impairments in part-whole representations of objects in two cases of integrative visual agnosia. *Cogn Neuropsychol*. 24:701–730.
- Biederman I. 1987. Recognition-by-components: a theory of human image understanding. *Psychol Rev*. 94:115–147.
- Biederman I, Bar M. 1999. One-shot viewpoint invariance in matching novel objects. *Vision Res*. 39:2885–2899.
- Bullier J. 2001. Integrated model of visual processing. *Brain Res Rev*. 36:96–107.
- Bussey TJ, Saksida LM, Murray EA. 2002. Perirhinal cortex resolves feature ambiguity in complex visual discriminations. *Eur J Neurosci*. 15:365–374.
- Coutanche MN, Thompson-Schill SL. 2015. Creating concepts from converging features in human cortex. *Cereb Cortex*. 25:2584–2593.
- Denys K, Vanduffel W, Fize D, Nelissen K, Peuskens H, Van Essen D, Orban G. 2004. The processing of visual shape in the cerebral cortex of human and nonhuman primates: a functional magnetic resonance imaging study. *J Neurosci*. 24:2551–2565.
- Desimone R, Albright TD, Gross CG, Bruce C. 1984. Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci*. 4:2051–2062.
- Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention. *Annu Rev Neurosci*. 18:193–222.
- Desimone R, Ungerleider LG. 1989. Neural mechanisms of visual processing in monkeys. In: Boller F, Grafman J, editors. *Handbook of neuropsychology*. New York: Elsevier Science. Vol. 2, p. 267–299.
- Devlin JT, Price CJ. 2007. Perirhinal contributions to human visual perception. *Curr Biol*. 17:1484–1488.
- Dumoulin SO, Wandell BA. 2008. Population receptive field estimates in human visual cortex. *Neuroimage*. 39:647–660.
- Eckhorn R. 1999. Neural mechanisms of visual feature binding investigated with microelectrodes and models. *Vis Cogn*. 6:231–265.
- Frank SM, Reavis EA, Tse PU, Greenlee MW. 2014. Neural mechanisms of feature conjunction learning: enduring changes in occipital cortex after a week of training. *Hum Brain Mapp*. 35:1201–1211.
- Grill-Spector K, Kourtzi Z, Kanwisher N. 2001. The lateral occipital complex and its role in object recognition. *Vision Res*. 41:1409–1422.
- Gross CG. 1992. Representation of visual stimuli in inferior temporal cortex. *Philos Trans R Soc Lond B Biol Sci*. 335:3–10.
- Gross CG. 2008. Single neuron studies of inferior temporal cortex. *Neuropsychologia*. 46:841–852.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*. 293:2425–2430.
- Hochstein S, Ahissar M. 2002. View from the top: hierarchies and reverse hierarchies review. *Neuron*. 36:791–804.
- Hubel DH, Wiesel TN. 1965. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J Neurophysiol*. 28:229–289.
- Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. *Nat Neurosci*. 8:679–685.
- Kastner S, De Weerd P, Pinsk MA, Elizondo MI, Desimone R, Ungerleider LG. 2001. Modulation of sensory suppression: implications for receptive field sizes in the human visual cortex. *J Neurophysiol*. 86:1398–1411.
- Kay KN, Rokem A, Winawer J, Dougherty RF, Wandell BA. 2013. GLMdenoise: a fast, automated technique for denoising task-based fMRI data. *Front Neurosci*. 7:1–15.
- Konen CS, Behrmann M, Nishimura M, Kastner S. 2011. The functional neuroanatomy of object agnosia: a case study. *Neuron*. 71:49–60.
- Kornblith S, Cheng X, Ohayon S, Tsao DY. 2013. A network for scene processing in the macaque temporal lobe. *Neuron*. 79:766–781.
- Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *Proc Natl Acad Sci USA*. 103:3863–3868.
- Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA. 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*. 60:1126–1141.
- Lafer-Sousa R, Conway BR. 2013. Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nat Neurosci*. 16:1870–1878.
- Lamme VA, Roelfsema PR. 2000. The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci*. 23:571–579.
- Lee ACH, Rudebeck SR. 2010. Human medial temporal lobe damage can disrupt the perception of single objects. *J Neurosci*. 30:6588–6594.
- Linke AC, Vicente-Grabovetsky A, Cusack R. 2011. Stimulus-specific suppression preserves information in auditory short-term memory. *Proc Natl Acad Sci USA*. 108:12961–12966.
- Liu J, Harris A, Kanwisher N. 2010. Perception of face parts and face configurations: an fMRI study. *J Cogn Neurosci*. 22:203–211.
- Macevoy SP, Epstein RA. 2009. Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr Biol*. 19:943–947.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB. 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci USA*. 92:8135–8139.
- McTighe SM, Cowell RA, Winters BD, Bussey TJ, Saksida LM. 2010. Paradoxical false memory for objects after brain damage. *Science*. 330:1408–1410.
- Mullen B. 1989. *Advanced basic meta-analysis*. Hillsdale, NJ: Erlbaum.
- Murray EA, Bussey TJ, Saksida LM. 2007. Visual perception and memory: a new view of medial temporal lobe function in primates and rodents. *Annu Rev Neurosci*. 30:99–122.
- Nakamura K, Matsumoto K, Mikami A, Kubota K. 1994. Visual response properties of single neurons in the temporal pole of behaving monkeys. *J Neurophysiol*. 71:1206–1221.
- Peissig JJ, Tarr MJ. 2007. Visual object recognition: do we know more now than we did 20 years ago? *Annu Rev Psychol*. 58:75–96.
- Peterson MA, Cacciamani L, Barense MD, Scalf PE. 2012. The perirhinal cortex modulates V2 activity in response to the

- agreement between part familiarity and configuration familiarity. *Hippocampus*. 22:1965–1977.
- 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.
- Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I. 2005. Invariant visual representation by single neurons in the human brain. *Nature*. 435:1102–1107.
- Reddy L, Kanwisher N. 2007. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Curr Biol*. 17:2067–2072.
- Reddy L, Kanwisher NG, Vanrullen R. 2009. Attention and biased competition in multi-voxel. *Proc Natl Acad Sci USA*. 106:21447–21452.
- Riesenhuber M, Poggio T. 1999. Hierarchical models of object recognition in cortex. *Nat Neurosci*. 2:1019–1025.
- Rosenthal R. 1994. Parametric measures of effect size. *The handbook of research synthesis*. pp. 231–244.
- Rottschy C, Eickhoff SB, Schleicher A, Mohlberg H, Kujovic M, Zilles K, Amunts K. 2007. Ventral visual cortex in humans: cytoarchitectonic mapping of two extrastriate areas. *Hum Brain Mapp*. 28:1045–1059.
- Sawamura H, Georgieva S, Vogels R, Vanduffel W, Orban GA. 2005. Using functional magnetic resonance imaging to assess adaptation and size invariance of shape processing by humans and monkeys. *J Neurosci*. 25:4294–4306.
- Singer W, Gray CM. 1995. Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci*. 18:555–586.
- Squire LR, Zola-Morgan M. 2011. The cognitive neuroscience of human memory since H.M. *Annu Rev Neurosci*. 34:259–290.
- Sripati AP, Olson CR. 2010. Responses to compound objects in monkey inferotemporal cortex: the whole is equal to the sum of the discrete parts. *J Neurosci*. 30:7948–7960.
- Tanaka K. 1996. Inferotemporal cortex and object vision. *Annu Rev Neurosci*. 19:109–139.
- Thorpe S, Fize D, Marlot C. 1996. Speed of processing in the human visual system. *Nature*. 381:520–522.
- Uhlhaas PJ, Pipa G, Lima B, Melloni L, Neuenschwander S, Nikolic D, Lin S. 2009. Neural synchrony in cortical networks: history, concept and current status. *Front Integr Neurosci*. 3:1–19.
- Ungerleider LG, Haxby JV. 1994. “What” and “where” in the brain. *Curr Opin Neurobiol*. 4:157–165.
- Vuilleumier P, Henson RN, Driver J, Dolan RJ. 2002. Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat Neurosci*. 5:491–499.
- Watson HC, Wilding EL, Graham KS. 2012. A role for perirhinal cortex in memory for novel object-context associations. *J Neurosci*. 32:4473–4481.
- Winters BD, Forwood SE, Cowell RA, Saksida LM, Bussey TJ. 2004. Double dissociation between the effects of peri-postrhinal cortex and hippocampal lesions on tests of object recognition and spatial memory: heterogeneity of function within the temporal lobe. *J Neurosci*. 24:5901–5908.
- Zeithamova D, Dominick AL, Preston AR. 2012. Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*. 75:168–179.