Cerebral Cortex, 2020;00: 1–19

ORIGINAL ARTICLE

OXFORD

Experience Transforms Conjunctive Object Representations: Neural Evidence for Unitization After Visual Expertise

Jackson C. Liang^{1,*}, Jonathan Erez², Felicia Zhang³, Rhodri Cusack⁴ and Morgan D. Barense^{1,5}

¹Department of Psychology, University of Toronto, Toronto, ON M5S 3G3, Canada, ²Department of Psychology, Brain and Mind Institute, Western Interdisciplinary Research Building, Western University, London, ON N6A 5B7, Canada, ³Department of Psychology, Princeton University, Princeton, NJ 08540, USA, ⁴School of Psychology, Trinity College Dublin, Dublin, Ireland amd ⁵Rotman Research Institute, Toronto, ON M6A 2E1, Canada

Address correspondence to Jackson C. Liang, Department of Psychology, University of Toronto, 100 St. George Street, Toronto, ON M5S 3G3, Canada. Email: jacksoncliang@gmail.com

Jackson C. Liang and Jonathan Erez contributed equally to this work

Abstract

Certain transformations must occur within the brain to allow rapid processing of familiar experiences. Complex objects are thought to become unitized, whereby multifeature conjunctions are retrieved as rapidly as a single feature. Behavioral studies strongly support unitization theory, but a compelling neural mechanism is lacking. Here, we examined how unitization transforms conjunctive representations to become more "feature-like" by recruiting posterior regions of the ventral visual stream (VVS) whose architecture is specialized for processing single features. We used functional magnetic resonance imaging to scan humans before and after visual training with novel objects. We implemented a novel multivoxel pattern analysis to measure a conjunctive code, which represented a conjunction of object features above and beyond the sum of the parts. Importantly, a multivoxel searchlight showed that the strength of conjunctive coding in posterior VVS increased posttraining. Furthermore, multidimensional scaling revealed representational separation at the level of individual features in parallel to the changes at the level of feature conjunctions. Finally, functional connectivity between anterior and posterior VVS was higher for novel objects than for trained objects, consistent with early involvement of anterior VVS in unitizing feature conjunctions in response to novelty. These data demonstrate that the brain implements unitization as a mechanism to refine complex object representations over the course of multiple learning experiences.

Key words: conjunctive processing, object representation, perceptual learning, perirhinal cortex, ventral visual stream

Introduction

Our environment is filled with objects that vary greatly in their visual complexity. Although the task of discriminating between objects can often be solved on the basis of single diagnostic features, more complex objects challenge the visual system to discriminate between the combinations of multiple features simultaneously. Given that much of the early visual system is dedicated to processing simple features within a narrow foveal receptive field, an important open question is how the brain supports rapid decisions about objects with spatially distributed visual features. One suggested mechanism is encapsulated by hierarchical models of the ventral visual stream (VVS) (Desimone and Ungerleider 1989; Gross 1992; Tanaka 1996; Riesenhuber and Poggio 1999), where representations are organized in increasingly complex conjunctions as information flows from posterior to anterior VVS. A recent theory, termed

© The Author(s) 2020. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Identifying Conjunctive Representations



Figure 1. Conjunction contrast. Our critical contrast investigated whether brain patterns of activity demonstrated explicit conjunctive coding (i.e., was the whole different from the sum of its parts?). To this end, we measured the 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 three different pairs of objects (i.e., A + BC, B + AC, and C + AB), which were matched in terms of their individual features (A, B, and C), but different in terms of their conjunction (i.e., AB, BC, and AC). 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.

the representational-hierarchical model, proposes that posterior VVS represents simple features, whereas perirhinal cortex (PRC) sits at the apex of this information stream, representing complex conjunctions at the level of a whole object (e.g., Bussey et al. 2002; Bartko et al. 2007; Cowell et al. 2010; Barense et al. 2010, 2012a; Lee and Rudebeck 2010; Peterson et al. 2012; Cacciamani et al. 2017). We recently developed a multivariate method that revealed an explicit conjunctive code in PRC (Erez et al. 2016). That is, neural patterns of PRC activity represented information specific to the conjunction of features comprising a complex object, over and above information regarding the features themselves (Fig. 1). However, we still do not understand how conjunctive representations in the brain might guide subsequent behavior.

One crucial insight revealed through everyday experience is that certain feature conjunctions can become highly familiar. For example, a favorite pet dog's individual features might be shared with countless other dogs. However, it is that pet's specific conjunction of ears, tail, and fur that has a special status, enabling it to stand out in a crowd of other dogs at the park. In the brain, a perceptual learning mechanism called unitization is thought to support the familiarization of feature conjunctions (Goldstone 1998; Goldstone 2000). According to this view, through extended learning, an object's individual features are integrated into a singular unitized representation. Behavioral evidence for unitization exists across many domains, including words (e.g., Smith and Haviland 1972; Laberge 1973), faces (e.g., Valentine 1988; Tanaka and Farah 1993), and novel visual objects (e.g., Gauthier and Tarr 1997). However, evidence linking these observed effects to any brain mechanism is limited to the observation that following training, neurons in the inferotemporal (IT) cortex shift their response selectivity towards the specific combinations of features in simple stimuli (Baker et al. 2002). On the one hand, a straightforward prediction would likewise suggest that repeated experiences drive conjunction-sensitive neuron populations in PRC to become more selective for their preferred conjunctions, and show attenuated responses toward the component features. On the other hand, the representational-hierarchical framework presents a compelling complementary organization for unitized conjunctions. Specifically, unitized conjunctive objects elicit a simpler representational scheme akin to those of more basic features supported by posterior regions of the VVS. The findings of Baker et al. (2002) are also consistent with this idea, by demonstrating that the representational architecture of primate IT cortex was originally sensitive to parts but could be repurposed to be sensitive to object wholes through learning. Under this view, PRC may guide the formation of these new unitized representations via feedback to posterior VVS. Such interactions are predicted by anatomical and functional studies (Clavagnier et al. 2004; Peterson et al. 2012; Cacciamani et al. 2017).

The goal of the present study was to identify the neural changes that support learning of object feature conjunctions. We scanned participants using MRI both before and after intensive training with a paradigm designed to promote the unitization of objects (Czerwinski et al. 1992). We assessed the impact of experience on conjunctive representations in the VVS, employing a unique design that allowed for the comparison of objects that were identical at the feature-level, yet different in terms of the conjunctive arrangement of these features. We used pattern information analysis to observe learning-related changes in PRC and posterior VVS at the conjunctive- and feature-levels. Finally, we used functional connectivity (FC) analysis to assess whether PRC shaped unitized representations through feedback to early visual areas.

Materials and Methods

Participants

Twenty right-handed adults with normal or corrected-to-normal vision (12 females, mean age = 23.6 years) were recruited from the University of Toronto community for Session 1 of the functional magnetic resonance imaging (fMRI) experiment. The results from fMRI Session 1 were originally reported in Erez et al. (2016). Fourteen participants returned to complete Visual Search Training and fMRI Session 2 and entered into a repeated-measures comparison of the pretraining Session 1 and posttraining Session 2. Additionally, we recruited four participants who only completed Visual Search Training and fMRI Session 2 but not fMRI Session 1. Thus, for analyses requiring only fMRI Session 2, 18 right-handed adults with normal or corrected-to-normal vision were available (11 females, mean age = 23.4 years). All neuroimaging participants provided written consent approved by the Baycrest Hospital Research Ethics Board and received monetary compensation for their participation. Thirty right-handed adults with normal or corrected-to-normal vision (19 females, mean age = 20.6) were recruited from the University of Toronto community for a control experiment to assess the effect of body color on visual processing. All participants in the control experiment were provided written consent approved by the University of Toronto Research Ethics Board and received monetary compensation for their participation.

Stimuli and Counterbalancing

Stimuli were color images of artificial 3D objects. These stimuli were constructed in Strata Design 3D CX 6 and were used in a previously published study of conjunctive representation in the visual system (Erez et al. 2016). Each object belonged to a specific family of seven objects whose members shared three distinct features (e.g., features A, B, and C; Fig. 2A) and body color. Each object within a family comprised 1, 2, or all 3 features such that, for example, a full set of seven objects in a given object Family could be summarized: A, B, C, BC, AC, AB, and ABC. Every object was created using two different viewpoints (as in Erez et al. 2016). However, viewpoint was not a factor of interest in this experiment and subsequent analyses were collapsed across viewpoint. Objects were presented centrally on the screen and had a visual angle of 5.1° × 5.3°.

Three groups of objects corresponding to specific body colors were viewed over the course of the experiment. The first group comprised two families of blue-bodied objects (Families 1 and 2; Fig. 2A) which were initially encountered in fMRI Session 1. One family (e.g., Family 1 or 2) was assigned to the "Session 1 To-Be-Trained" condition, and the other family to the "Session 1 Novel" condition. These labels refer not to how much visual exposure they received during Session 1 but rather their status relative to the objects viewed in fMRI Session 2. Specifically, the "To-Be-Trained" objects were held over to be used as Trained objects in fMRI Session 2, but the "Novel" objects were replaced by a new set of objects in fMRI Session 2 (e.g., Family 5 or 6; Fig. 2C). Hereafter, Families 1 and 2 will be referred to as Session 1 To-Be-Trained and Session 1 Novel objects, respectively, when describing neuroimaging analyses from Session 1, reflecting their still-equivalent training status at that point. When describing neuroimaging analyses from Session 2, the held-over Family 1 or 2 will be referred to as Session 2 Trained objects, whereas Families 5 and 6 will be referred to as Session 2 Novel objects, reflecting their actual differences in training status at that point. During Visual Search Training, the bluebodied families were studied extensively as part of the Learning Set of objects. A group of objects comprising two families of yellow-bodied objects (Families 3 and 4; Fig. 2B) belonged to the Novel Probe Set and were encountered only on the sixth and final day of Visual Search Training. The Novel Probe Set allowed us to assess behavioral performance on a new set of objects with no training history.

fMRI Session 1

To obtain a pretraining measure of the brain responses to novel objects, we scanned participants while they performed four runs of a 1-back repetition detection task with all of the objects from Families 1 and 2 (Fig. 2A). On each trial, an object was presented on screen for 1 s followed by a 2 s interstimulus interval (ISI). Participants were instructed to rapidly respond with a button press whenever an object appeared that was a duplicate (regardless of viewpoint) of the preceding one. The trial order of presented objects was organized in blocked fashion, where all members of one family across both viewpoints were viewed together in a mini-block of 14 objects. Each block contained two mini-blocks of the same family (totaling 28 trials). The trial order of objects within each mini-block was calculated using OptSeq2 (Dale 1999), to ensure that the neural response to each individual object could be reliably estimated using a general linear model (GLM). Three blocks each of Family 1 and Family 2 alternated, totaling six blocks per run. Whether each run began with a block of Family 1 or Family 2 was counterbalanced across participants. Each block contained 1–4 targets (i.e., sequential repeats) such that the proportion of target trials across the experiment was 10%. Each run lasted 11 min 30s, and for every 42 s of task time, there was an 8-s break during which a fixation cross was presented on screen. "Functional localizer". To obtain functional regions-of-interest (ROIs) representing distinct category-specific components of the VVS, we conducted a scanned localizer task where participants viewed face, object, scrambled object, and scene stimuli while engaged in a 1-back repetition detection task (for localizer details, see Erez et al. 2016).

Visual Search Training (6-Day Behavioral Protocol)

We invited participants to return for a 6-day Visual Search Training protocol designed to promote the unitization of object features (Czerwinski et al. 1992). On this task, unitization is expected to improve conjunctive search in general, but the benefit is most apparent in the presence of many distractors. Thus, the evidence of unitization can be identified by the observation of shallower search slopes following visual training (Fig. 3A), reflecting minimal search time cost despite the presence of more distractors (e.g., Ashbridge et al. 1997).

Participant recruitment followed a long delay (minimum 100 days, average 287 days), during which the longitudinal components of this study (Visual Search Training and fMRI Session 2) were designed and participants from fMRI Session 1 were recruited. Importantly, there was always a 1-day delay between the end of the Visual Search Training protocol and fMRI Session 2, ensuring that training effects were recent and consistent across participants. Participants completed six 1-h visual training sessions, with each session taking place on a different day. Each trial began with a display of a black fixation cross for 1 s at the center of the screen, followed by a display of a target object for 1 s, a blank screen for 1 s, and a search array comprising between 1 and 8 objects (Fig. 2B, right panel). These objects remained on the screen until a response was made. On each trial, participants were asked to indicate whether the target object was present or absent by pressing one of two green colored buttons on a laptop keyboard ("L" if the target was present and "A" if the target was absent). Participants were instructed to respond as quickly as possible on each trial, without sacrificing accuracy. One-second auditory feedback was provided for each trial in the form of a high/low tone for correct/incorrect response, respectively.

The stimuli used in Visual Search were the blue-bodied families from Session 1 (Families 1 and 2; Fig. 2B), comprising 14 objects each (seven from each viewpoint). Additionally, features from both families were recombined to form a new set of 24 bluebodied objects (12 from each viewpoint) to be used as distractors only. Features that occupied the same spatial location on the object were not recombined. Altogether, these objects made up the Learning Set of stimuli that would be extensively studied on all 6 days of visual training. On the last day of training, a Novel Probe Set of objects was introduced, comprising two new yellowbodied families (Families 3 and 4) of 14 objects each (seven from each viewpoint). Features from both of these families were also recombined to form a new set of 24 yellow-bodied objects (12 from each viewpoint) to be used as distractors only. The Novel Probe Set enabled the observation of behavioral performance on a set of objects with no training history, in direct contrast to the Learning Set.



Figure 2. Experimental design. (A) First fMRI session. The stimulus set comprised two families of blue-bodied objects composed of a single-feature, double-features, or three features. One of these families was designated "To-Be-Trained", to be viewed again during Visual Search Training and fMRI Session 2. The other family was designated the "Novel" Set, to be viewed again during Visual Search Training but not in fMRI Session 2 (this designation was counterbalanced across participants). Participants performed a 1-back repetition detection task. Repetitions included objects that repeated exactly or objects that were repeated but viewed from a different angle (red-bordered squares). (B) Visual Search Training. The stimulus set comprised all objects viewed during the first fMRI session (Learning Set), along with two new families of yellow-bodied objects used exclusively on Day 6 of training (Novel Probe Set). Features from Families 1 and 2 were recombined to generate distractor objects for blue-bodied search traigets, while Families 3 and 4 were used to create yellow-bodied distractor objects. In each visual search trial, participants indicated whether the sample stimulus was present within an array of 1–8 objects. (C) Second fMRI session. The organization of stimuli was similar to the first fMRI session. Critically, the Session 1 To-Be-Trained objects (Family 1 or 2) were reused as the Session 2 Trained object set, while a new family of red-bodied objects was introduced here as the Session 2 Novel object set (Family 5 or 6).

On each trial, a target object was randomly selected from the Learning Set. Half the trials were "target present" trials. A target was considered present if the same object appeared in the search array from either the same or a different viewpoint. The remaining half of trials were "target absent" trials, in which only distractor objects were presented in the search array. Possible distractor objects were chosen to overlap with the target object either in terms of the number of features present or in terms of



Figure 3. Visual search training. (A) To obtain a measure of the behavioral improvement expected to result from unitization, we calculated the slope of the reaction time increase from smaller to larger visual search set sizes. Unitized representations have a shallow search slope, mitigating the usual increase in reaction time when searching in larger visual arrays. (B) Search slopes decreased from the first to the last day of training (gray dots), suggesting that some degree of unitization had occurred. On the last day, search slopes were shallower for trained objects relative to a new set of novel objects (white dot), allowing us to eliminate the possibility that the improvements were entirely related to general practice effects. $*P \le 0.05$. $**P \le 0.01$. Error bars represent the standard error of the mean across participants.

the specific features present. For example, if the target was an object containing a single feature A, then distractors included all other single feature objects and all objects containing feature A. For two- and three-featured objects (e.g., A_1B_1), possible distractors included any objects that contained one or more of its component features (e.g., A_1 , B_1 , A_1C_1 , B_1C_1 , $A_1B_1C_1$, A_1B_2 , A_1C_2 , A_2B_1 , B_1C_2 , and $A_1B_1C_2$). Distractors always shared body color with the target. Objects in the search array were presented at random locations on an imaginary circle that included 10 possible slots around the (now absent) target object.

In each training session, participants completed four blocks of the visual search task, each lasting approximately 15 min. Each block comprised 20 trials from each of 1–8 possible stimuli set sizes, resulting in (20 trials \times 8 set sizes) 160 trials per block or (160 trials \times 4 runs) 640 trials per training session. On the sixth and last training session, participants completed two blocks with the blue-bodied Learning Set, followed by two training blocks with the yellow-bodied Novel Probe set they had not seen before.

fMRI Session 2

On the day following the completion of visual training, participants returned for the second scanning session. To obtain a posttraining measure of the difference in brain response to learned and novel objects, participants were scanned while performing four runs of the 1-back repetition detection task with one family of blue-bodied objects representing the Session 2 Trained condition (e.g., Family 1 or 2; Fig. 2A) and one family of red-bodied objects representing the Session 2 Novel condition (e.g., Family 5 or 6; Fig. 2C). Specifically, to ensure that observed effects were general across Session 2 Trained and Session 2 Novel families, half of the participants viewed one set of Session 2 Trained objects (Family 1) along with Session 2 Novel objects (Family 5), while the remaining participants viewed another set of Session 2 Trained objects (Family 2) along with the other Session 2 Novel objects (Family 6). The organization and timing of trials and target distribution in Session 2 mirrored Session 1, with each run alternating between the Trained and Novel conditions. Functional localizer. We conducted a second localizer in similar fashion to the first scanning session, but with a new set of face, object, scrambled object, and scene stimuli.

fMRI Acquisition

Imaging data were acquired using a 32-channel head-coil on a 3.0-T Siemens MAGNETOM Trio MRI scanner housed at the Rotman Research Institute at Baycrest Hospital in Toronto, ON. Each scanning session began with a wholebrain T₁-weighted magnetization-prepared rapid gradientecho (MPRAGE) structural image acquired in 160 1-mm-thick oblique axial slices (192 × 256 inplane matrix). Functional brain images were then acquired using a T_{2*}-weighted echo-planar imaging (EPI) sequence [time repetition (TR) = 2000 ms; time echo = 30 ms; flip angle = 78° ; field of view = 20 cm; interslice gap = 0.5 mm; and 3.1 \times 3.1 \times 2.0 mm resolution]. A total of 389 functional volumes were acquired for each participant over four scanning runs in Session 1. The second session was performed using an identical scanning protocol, resulting in 778 functional volumes acquired across the two experimental sessions. To allow for T₁ stabilization, the first trial of each run did not commence until 10 s (5 TRs) after the beginning of the functional acquisition; data from this initial 10-s time period were discarded and not used in subsequent analyses.

Imaging Data Preprocessing

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 (Cusack et al. 2015). In preparation for multivoxel pattern analysis (MVPA) and multidimensional scaling (MDS) analysis, the data were preprocessed, which included the realignment of the data to the first functional scan of each run (after 5 dummy scans were discarded to allow for signal equilibrium), slicetiming correction, co-registration of functional and structural images, nonlinear normalization to the Montreal Neurological Institute (MNI) template brain, and segmentation of gray and white matter. Time series from each voxel were high-pass filtered with a 128-s cut-off. The data were then "denoised" by deriving regressors from voxels unrelated to the experimental paradigm and entering these regressors in a 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

blood oxygen level-dependent (BOLD) response amplitudes (beta weights). To avoid circularity, this design matrix for denoising component selection 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 R2 less than 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).

ROI Definitions

In preparation for the MDS and FC analyses (Analyses 2 and 3), we defined five ROIs a priori. Two ROIs, V1, and PRC were defined anatomically. Specifically, V1 was demarcated using FreeSurfer, which utilized landmarks localized to specific coordinates in a cortical surface model as demonstrated in Hinds et al. (2008). PRC was defined by an anatomical probability map created by Devlin and Price (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). Three were functionally defined regions well-established as part of the VVS: lateral occipital complex (LOC), fusiform face area (FFA), and the parahippocampal place area (PPA). 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), 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 LOC and PPA ROIs were bilateral, except for one participant in whom the right LOC could not be localized. FFA ROIS were bilateral in 10 participants; for six participants, the left FFA could not be localized, and for one participant, the right FFA was not localized.

In addition, we defined one ROI post hoc. This ROI, hereafter referred to as the "Trained > Novel Conjunction Contrast ROI", was functionally defined and captured the regions of posterior VVS that showed the greatest training-related increase in conjunctive coding. For each participant, we generated a 10-mm radius sphere around the peak voxel in posterior VVS for the Trained > Novel Conjunction Contrast searchlight from fMRI Analysis 1. Because this ROI was functionally defined by fMRI Analysis 1, it was only interrogated in fMRI Analyses 2 and 3 and in our Control Analysis 2 to rule out nonlinear biases in fMRI Analysis 1 (described below).

fMRI Analysis 1: Identifying Training-related Changes in Conjunctive Object Representations Using MVPA

We predicted that the spatial distribution of conjunction-based neural responses would change following learning such that posterior VVS regions would be recruited to represent the

unitized conjunctions. 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 modified by training. We examined the patterns of brain activity evoked by three features distributed across two individually presented objects during a 1-back task (Fig. 2C). Our critical "Conjunction 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. 1). 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. 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. 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.

Importantly, explicit conjunctive coding has been minimally studied relative to other prevailing coding schemes, such as feature-coding (Eckhorn 1999; Singer and Gray 1995). A recent human fMRI experiment found that the additive multivoxel patterns of individual objects (e.g., A and B) could largely explain the pattern evoked by complex scene-like conjunctions comprising A and B together (MacEvoy and Epstein 2011). Such demonstrations that the whole can share much in common with the sum of its parts have been taken as evidence that feature-coding can account for all complex object representation. However, explicit conjunctive coding proposes that certain brain regions are tuned specifically to conjunctions rather than the component features. In the past, the ability of univariate fMRI to localize this type of representational scheme within the human brain has been limited by the difficulty of deconfounding the response to feature conjunctions from the response to the individual features. Recently, the multivariate "conjunction contrast" method developed by Erez et al. (2016) revealed explicit conjunctive coding in PRC and LOC, giving us the first insight into a "whole" object representation above and beyond the sum of its parts. Using the procedure described below, we implement and extend this approach to assess how the conjunctive code in VVS changes as a function of experience.

The conjunction-based representation for novel objects was previously shown by Erez et al. (2016) using the fMRI Session 1 data, where all objects were newly introduced. Thus, the principal goals of the current analysis were 2-fold: (1) to replicate this original finding with the Novel objects from fMRI Session 2 and (2) to investigate whether unitization training would evoke a more posteriorly distributed response for the Session 2 Trained objects. To this end, all 18 participants who were scanned in both fMRI sessions were entered into a MVPA of their Session 2 data. Following the procedure established by Erez et al., within each run, we identified the onsets for each individual object (e.g., A, B, C, AB, BC, and AC) split by training status (Session 2 Trained objects and Session 2 Novel objects) and viewpoint (1 and 2). To allow the comparison of the similarity of patterns evoked by multiple repetitions of the same object, these object conditions were additionally split into two subdivisions whose onsets were relatively equidistant in time; one subdivision comprised the odd trials of every object, and the other subdivisions comprised the even trials. Our model then specified a single regressor for each unique conjunctive pair of objects (i.e., A + BC, B + AC, and C + AB; Fig. 1). This was done separately for each condition split by training status, viewpoint, and subdivision. For example, all odd trial events (subdivision 1) corresponding to the singly presented "A" object from viewpoint 2 of the Session 2 Novel object set were concatenated with all odd trial events corresponding to the singly presented "BC" object from viewpoint 2 of the Session 2 Novel object set to create the single regressor for the Session 2 Novel "A+BC" conjunction from viewpoint 2 in subdivision 1. This procedure resulted in 24 regressors of interest per run [2 (training statuses) \times 2 (viewpoints) \times 3 (conjunctions) \times 2 (subdivisions)]. We also specified an additional 8 regressors of no interest for each run: trials containing singly presented threefeature objects ("ABC"), trials in which participants responded with a button press on the 1-back task, and six realignment parameters to remove residual noise due to motion. Events were modeled with a delta (stick) function corresponding to the stimulus presentation onset convolved with the canonical HRF as defined by SPM8, and entered into GLMs for each run resulting in parameter estimates (β) indexing the magnitude of response during each of 8 repetitions (4 runs \times 2 subdivisions) of a given conjunction.

MVPA was performed on these parameter estimates using Pearson correlation (for an introduction to the technique, see Kriegeskorte et al. 2008). First, the correlations between every combination of those repetitions were calculated and averaged to obtain the within-conjunction correlation measure for a given conjunction (e.g., Fig. 4C, on-diagonal squares). This calculation was repeated for every conjunction derived from the Session 2 Trained and Session 2 Novel families across both viewpoints, totalling 12 conjunction conditions (3 conjunctions × 2 families \times 2 viewpoints). Between-conjunction correlations (e.g., Fig. 4C, off-diagonal squares) were computed similarly, being the average correlation between the eight repetitions of one conjunction condition with the eight repetitions of any other conjunction condition. This calculation was repeated for all possible pairings of the 12 conjunction conditions to obtain the final 12 \times 12 contrast matrix (similar to Linke et al. 2011).

A searchlight analysis was performed to test for multivoxel pattern information across the brain (Kriegeskorte et al. 2006). 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. Within this ROI, our predefined contrast matrix 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 (Fig. 4, left panels). First, to test for the replication of the findings from Erez et al. (2016), we localized the conjunctive code for Session 2 Novel objects by subjecting the contrast comparing within and between conjunction types to a two-sample t-test for Session 2 Novel objects only (Fig. 4A, left). Then, we tested the prediction that the conjunctive code is modified through experience, by subjecting the contrast comparing within and between conjunction types to a two-sample t-test for Session 2 Trained objects only (Fig. 4B, left). Critically, to directly capture the distribution of regions where training status tracked changes in the conjunctive code, we tested for regions that demonstrated a stronger conjunctive code for Session 2 Trained objects relative to Session 2 Novel objects (Fig. 4C, left). This analysis was first 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. These single-subject t-maps were then 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 1-sample t-test relative to zero). Results shown are overlaid on the single-subject MNI brain template. Experimental effects exceeding a threshold of P < 0.001 (uncorrected) were considered significant. Cluster-level significance was obtained via Monte Carlo simulation implemented in AlphaSim. All reported clusters exceeded a threshold of P < 0.001 corrected by Family-wise error (FWE).

fMRI Analysis 2: Visualizing the Separation of Individual Object Representations Using MDS Analyses

We predicted that extended experience with an object would alter the perception of its visual form such that the representations of individual objects would become more distinct with increased familiarity. In addition, the representationalhierarchical view proposes a layered model that enables the learning-related separation of individual object features (i.e., feature tuning) within posterior VVS and the separation of feature conjunctions within anterior VVS. Crucially, experience with features solely enhances feature tuning; to enhance tuning of conjunctive representations comprising those features, the specific conjunctions must themselves be repeatedly experienced (Sadil and Cowell 2016). Thus, we used an ROI-based approach with MDS analyses on the data from both pre and posttraining fMRI Sessions to ask which brain regions demonstrated greater distinction between object representations after training. The 14 participants who completed both fMRI Sessions 1 and 2 were entered into the MDS analysis.

To obtain the brain response evoked by individual objects, we modeled trial events in a similar fashion to fMRI analysis 1. Within each run, we specified the onsets for each individual object (e.g., A, B, C, AB, BC, AC, and ABC) split by training status (Session 1 To-Be-Trained objects and Session 1 Novel objects), viewpoint (1 and 2), and subdivisions (1 and 2). This resulted in 56 regressors of interest per run [2 (training statuses) \times 2 (viewpoints) \times 7 (conjunctions) \times 2 (subdivisions)]. We also modeled seven regressors of no interest for each run: trials in which participants responded with a button press on the 1-back task and six realignment parameters to remove residual noise due to motion. Events were modeled with a delta (stick) function corresponding to the stimulus presentation onset convolved with the canonical HRF as defined by SPM8, and entered into GLMs for each run resulting in parameter estimates (β) indexing the magnitude of response during each of eight repetitions (4 runs \times 2 subdivisions) of a given object. These GLMs were conducted for each run in fMRI Sessions 1 and 2.

We employed a priori ROIs that each embodied different layers of the representational hierarchy: V1 representing the simplest visual features, FFA, LOC, and PPA representing more



Figure 4. Training-related changes in the conjunction contrast for (A) Session 2 Novel objects and (B) Session 2 Trained objects. Left: MVPA correlations within and between conjunctions were summarized in a matrix structure. This contrast tested whether correlations between the repetitions of the same conjunctions (dark squares) were greater than correlations between different conjunctions, despite comprising the same features (light squares). Right: regions where the conjunctive representations for objects were different from the sum of their individual features. The conjunction contrast for Novel objects evoked the focal clusters of conjunctive representation within right PRC and bilateral LOC, whereas the conjunction contrast for Trained objects evoked an extensive posterior representation within the VVS. Because our effects were primarily related to changes in the distribution and not necessarily the magnitude of significant clusters, a different view was selected for each conjunction contrast to best capture their extents. x, y, and z values indicate where each viewing plane was located. (C) Conjunction Contrast: Trained > Novel. We identified regions that showed greater conjunctive coding for Trained objects relative to Novel objects. All maps are shown (voxel-level *P* < 0.001, uncorrected, cluster-level *P* < 0.001 FWE). All clusters shown exceeded a cluster-level extent calculated from Monte Carlo simulation implemented in AlphaSim and thresholded at *P* = 0.001 FWE. The conjunction contrast for Trained > Novel revealed a focal cluster of conjunction representation within left PRC and a broader representation within posterior VVS.

complex features, and PRC representing the most complex feature conjunctions composing objects. We also employed the Trained > Novel Conjunction ROI to examine how individual objects are represented within areas showing the strongest conjunctive code. Within each ROI, we calculated the pattern similarity between GLM-based beta estimates corresponding to each individual object. The pattern similarity was expressed simply as the Pearson's correlation between the linearized voxel patterns extracted from an ROI mask of two different beta images. This process was repeated on multiple subsections of data for each participant; specifically, the beta patterns for every object shown in a given run and subdivision (e.g., Run 1, Subdivision 1) were correlated to generate a representational similarity matrix (RSM). Across four runs and two subdivisions, this process resulted in eight RSMs for each participant. In preparation for MDS, each participant's eight RSMs were averaged into a single RSM and converted into a representational dissimilarity matrix (RDM) by taking the difference between each matrix value and 1 (Fig. 5A). We implemented an MDS procedure that iteratively randomized exemplar positions in high-dimensional space, with error being minimized when the spatial distances between exemplar positions closely matched the target RDM. Individual squares of the RDM exhibiting the highest variance across participants (top 1%) were replaced with blanks, allowing them to act as additional degrees of freedom for error minimization on the remainder of the data. In particular, this means that even if the observed distance between any two exemplars is highly variable, their final positions can still be estimated on the strength of their more reliable relationships with other exemplars in the dataset. MDS was implemented in MATLAB using a nonmetric stress criterion for obtaining goodness of fit (i.e., STRESS1, normalized by the sum of squares of interpoint distances), and was conducted for every RDM from each ROI and participant.

We visualized MDS plots in 2D, where the distances between individual objects represent the corresponding dissimilarities between their neural responses (Fig. 5B). We then calculated the mean interexemplar distances between points within each session and training condition in each ROI. Because the Session 1 Novel objects and Session 2 Novel objects were represented by unrelated stimuli with different body colors (e.g., Family 2 vs. Family 5), repeated-measures ANOVA was not conducted to determine the contribution of training status or the interaction between training status and session as factors in the data. Instead, planned pairwise comparisons were calculated independently within each session to determine whether there were significant differences in interexemplar distances between objects from each training condition. To account for the contribution of stimulus differences (i.e., body color) to participants' ability to unitize in the experiment, we conducted an independent control experiment described in "Control Analysis 1." Additional pairwise comparisons assessed whether the mean interexemplar distance changed (1) between Session 1 To-Be-Trained objects and Session 2 Trained objects and (2) between Session 1 Novel objects and Session 2 Novel objects. Note that while 2D-MDS calculated from a group-averaged RDM was the most visually clear rendering of the trends we observed in our data as they pertain to our two factor design (session and training status), our inferential statistics were performed on the results from 3D-MDS calculated on a per-subject basis. This allowed us to ensure that the trends we observed in the visualized group MDS were consistently present across participants. In combination with our RDM outlier detection procedure, this makes MDS as a crucial component of our statistical approach.



Figure 5. Separation of individual object representations. (A) We calculated the pairwise pattern similarities between the brain responses evoked by each Novel object and To-Be-Trained object in Session 1, and Novel and Trained object in Session 2. Similarities were converted into dissimilarities and populated a matrix where darker squares represented higher dissimilarity. For illustrative purposes, only a portion of an example participant's full dissimilarity matrix is shown here. (B) A group-level dissimilarity matrix was calculated and then fitted using an iterative MDS function. The result was an MDS plot, where the spatial distance between plotted objects reflected the dissimilarity between their brain responses. (C) We calculated MDS plots for each ROI, with the Session 1 objects plotted in gold, and the Session 2 objects plotted in blue. To-Be-Trained and Trained (filled in) dots represent the same family of objects (e.g., Feature Set 1 from Fig. 1) presented before and after visual training, while the Novel (hollow) dots represent two different sets of objects (e.g., Feature Sets 2 and 5) introduced as unfamiliar objects in Session 1 and Session 2, respectively. We observed significantly greater distances between Session 2 objects relative to distances between Session 1 objects in FFA, LOC, PPA, and V1, but not PRC, We also observed a significant difference between Sessions in our Trained > Novel Conjunction Contrast ROI, representing a sphere around the peak of our Trained > Novel Conjunction Contrast searchlight from Figure 4. Thus, visual training was associated with greater distinction between the representations of individual objects in the VVS.

fMRI Analysis 3: Assessing Changes in Connectivity Between VVS Subregions Following Training

We considered that the learning of object conjunctions would be supported by close interactions between those brain regions supporting conjunctive representations and those supporting single-feature representations. Specifically, we hypothesized *a priori* that feedback from PRC played an important role in shaping object representations in posterior VVS during learning, based on established anatomical and functional studies in support of feedback influence between these regions (Clavagnier et al. 2004; Peterson et al. 2012; Barense et al. 2012b). Therefore, we limited our analyses to understanding how PRC was functionally connected to the other posterior VVS regions over the course of fMRI Sessions 1 and 2. The 14 participants who completed both fMRI Sessions 1 and 2 were entered into the FC analysis.

In the Session 1 data, we used PRC as a seed region to create separate time-series of FC between each ROI and the seed (Fig. 6A). Predicated on the assumption that FC may greatly fluctuate especially in the context of a rapid event-related design, we used a sliding window approach using a window size of four TRs (i.e., 8 s; see Hutchison et al. 2013 for a review of the technique). This window size was chosen to ensure that a minimum of one hemodynamic response could be captured by each window. The mean activation across voxels for each ROI (PRC, FFA, PPA, LOC, V1, and the Trained > Novel Conjunction Contrast ROI) was extracted from each TR to obtain time-series restricted to the sliding window. We then calculated Pearson's correlations between the windowed time-series of the PRC seed and all other ROIs. The resulting correlation metrics were then assigned to the first TR of the window. The window was then moved forward one TR (i.e., 2 s), and the process was repeated until the window terminated at the end of the session. We averaged the correlations across timepoints within each of the Session 1 To-Be-Trained and Session 1 Novel conditions to obtain separate connectivity measures for To-Be-Trained and Novel objects in Session 1. This process was repeated for the Session 2 Trained objects and Session 2 Novel objects. In Session 1, the familiarity status of all objects were the same (the "To-Be-Trained" and "Novel" labels indicated only whether they would subsequently be used as the Trained set in Session 2), and thus we predicted that we would find no FC differences between the Session 1 To-Be-Trained and Session 1 Novel objects. However, we expected that Visual Search Training would have a differential effect on Session 2 Trained objects relative to Session 2 Novel objects. Because the Session 1 Novel objects and Session 2 Novel objects were represented by unrelated stimuli with different body colors (e.g., Family 2 vs. Family 5), repeated-measures ANOVA was not conducted to determine the contribution of training status or the interaction between training status and session as factors in the data. Instead, planned pairwise comparisons were calculated independently within each session to determine whether there were significant differences in FC with PRC between objects from each training condition. To account for the contribution of stimulus differences (i.e., body color) to participants ability to unitize in the experiment, we conducted an independent control experiment described below.

Control Analysis 1: Perceptual- Versus Training-dependent Biases in Visual Processing

The neuroimaging dataset from the pretraining fMRI session (fMRI Session 1) was obtained from an earlier experiment

designed to identify explicit conjunctive codes within VVS (Erez et al. 2016). A crucial aspect of that approach was to ensure that each conjunction was drawn from one of two consistent and unique feature sets (Families 1 and 2, respectively; see Fig. 2A). This method naturally extended to our present goal of understanding how conjunctive coding is transformed via experience; in one counterbalancing group, we were able to select one of the two object Families for additional training (e.g., Family 1) and introduce a novel object Family (e.g., Family 5; see Fig. 2C) to identify any effects of training status in the posttraining fMRI session (fMRI Session 2). In the other counterbalancing group, we selected Family 2 for additional training and introduced novel object Family 6 for posttraining comparisons. To ensure that families would not be confused across sessions and that all components of the novel objects from Session 2 were entirely novel, Session 1 Families had blue bodies, whereas Session 2 bodies were red.

Our pretraining analysis approach was to compare two Families of equivalent training status (i.e., neither Family 1 nor Family 2 had undergone any visual training), leading to a comparison between two blue-bodied object Families. On the other hand, our posttraining analysis approach was to compare two Families of differing training status (for example, Family 1 underwent training, but Family 5 was novel), leading to a comparison between a blue-bodied object Family and a red-bodied object Family. It was, therefore, necessary to ensure that our neural effects presumed to arise from differences in training status could not be instead the result of explicit differences in visual processing performance between object body colors *per se*.

To that end, we conducted a control visual search study using an independent sample of 30 participants (described in "Participants" section above) and directly tested the possibility that an interaction between a Family's body color and its physical features led to explicit differences in behavioral performance that could bias the neuroimaging results. Specifically, we assessed participants' visual search performance on the three pairs of object Families corresponding to the Families observed in the main experiment: (1) the blue-bodied objects introduced in fMRI Session 1; (2) the yellow-bodied objects introduced during Visual Search Training; and (3) the red-bodied objects introduced in fMRI Session 2. Unlike the Visual Search Training phase in the main experiment, however, here, all objects were introduced within the same and only testing session. Furthermore, the order of presentation of the three colors was counterbalanced across participants, ensuring that any significant differences in behavioral performance between Families could only be the result of immediate differences in visual processing.

Although the overall structure of the control experiment was abbreviated relative to Visual Search Training in the main experiment (data were collected in a single day compared with six separate days), they were otherwise identical in terms of the structure of their individual trials. Specifically, each trial began with a fixation cross displayed for 1 s, followed by a sample period for 1 s wherein participants were instructed to hold the target object in mind. This was followed by a delay period of 1 s, and finally a visual search period wherein participants were instructed to indicate whether the target was present in a visual search array containing 1–8 distractor objects sharing the same body color. The construction of distractor objects and their placement within each search array were implemented in identical fashion to the Visual Search Training from the



Figure 6. Training-related changes in FC between PRC and posterior VVS. (A) We measured the moment-to-moment FC between PRC and posterior VVS ROIs using a sliding window that comprised four functional volumes (black box). We calculated the average FC values during blocks of Session 2 Novel objects (white background) and blocks of Session 2 Trained objects (gray background) to obtain FC measures for each training condition. (B) Following visual training, we observed significantly greater FC with PRC for Session 2 Novel objects relative to Session 2 Trained objects in PPA, V1, and the Trained > Novel Conjunction Contrast ROI. By comparison, we observed no significant differences in FC with PRC between Novel and Trained objects in any of the tested ROIs in Session 1. Thus, heightened interactions between PRC and VVS regions were uniquely associated with the presentation of novel objects, consistent with the idea that PRC guides the formation of unitized representations in posterior VVS.

main experiment (for full details, see Material and Methods for Visual Search Training). Altogether, there was one block of 208 trials with a blue-bodied target object, one block of 208 trials with a yellow-bodied target object, and one block of 208 trials with a red-bodied target object. Each block of a given body color was split across two consecutive testing runs of 104 trials each. Within a block, every object from each feature set (e.g., Families 1 and 2 for blue-bodied objects) and each viewpoint (1 and 2) were used as target objects four times. The specific order in which objects from within each Family and viewpoint were used as targets within a block was randomized across participants.

Because each body color was presented in consecutive task blocks, we anticipated that participants' visual search performance would improve as they transitioned from one color to the next. However, our goal was to determine the effect of body color on search performance above and beyond practice effects arising from the ordering of those colors across task blocks. Therefore, we disentangled the relative impact of order, color, and the interaction between order and color by calculating a linear mixed model with those factors entered as fixed effects. Participant was entered into the model as a random effect, using variance components as the covariance structure. We conducted one linear mixed model for each of our two dependent variables: hitrate and search slope. In our statistical reporting, we acknowledge that reporting effect sizes for each of the parameter estimates in the mixed model is subject to varying methodologies and interpretability issues. Therefore, we did not report effect sizes. Instead, we reported the F-statistic and P-value solely as a way to test for the presence of significant factors in our model and not as a way to assess the magnitude of their effects. Because we designed these objects to be similar in visual complexity, we did not expect to observe a significant effect of color. Nonetheless, for each dependent variable, we also conducted follow-up pairwise comparisons between each body color to ensure that no particular set of object Families was processed with greater speed or accuracy relative to any other. We calculated a paired t-test for each comparison and reported effect size as Cohen's d.

Control Analysis 2: Non-linear Coupling of Neuronal and BOLD Response

We also tested the possibility that signal saturation at the level of individual voxels for features could have inflated decodability in the multivariate patterns. Voxels already near saturation in response to any single-feature would likely respond more weakly than expected when a second feature is added, due to the saturation of the BOLD signal (i.e., they would show subadditivity, Supplementary Fig. 3A). This would introduce a difference in the responses to multifeature objects that could have been misinterpreted as a neuronal conjunctive response. To investigate the possibility of this effect being present in our dataset, we characterized the signal properties of the voxels in our Trained > Novel Conjunction Contrast ROI—the region of the brain demonstrating maximal response to our primary multivariate contrast of interest.

Voxel-wise Signal to Single- and Double-feature Objects

First, we examined how, at the level of individual voxels, the response scaled up from single-feature objects (e.g., A or B) to double-feature objects (e.g., AB). For signal saturation to occur, the responses to double-feature objects must be higher than that evoked by single-feature objects (Supplementary Fig. 3A). To investigate this, we overlaid the single- and double-feature distributions of voxel-wise response amplitudes, expressed as the proportion of total number of voxels in the ROI. To quantify whether any portion of the distributions significantly differed between single- and double-feature objects, we calculated paired Student's t-tests at each level of response amplitude. To ensure that any slight differences were not due to signal saturation, we tested for a biophysical ceiling effect in each single-feature and double-feature object response distribution. We exploited the fact that signal saturation would limit the highest levels of response amplitude and negatively skew the overall response distribution (Supplementary Fig. 3B, blue curve). To investigate this possibility, we calculated whether the doublefeature response distribution was more negatively skewed relative to the single-feature response distribution.

The Influence of Univariate Single-Feature Response Amplitude on Multivariate Contrasts

Another possibility is that a divisive normalization mechanism (Heeger 1992) may operate to prevent saturation when presented with multiple features (MacEvoy and Epstein 2009). This process could attenuate the responses to individual features when presented within a conjunction (e.g., AB) but not when presented singly (e.g., A), leading to a mismatch in the univariate response to that feature across different object conjunctions in our study (e.g., the response to A being lower in AB+C relative to A+BC). Such a mismatch could influence our multivariate contrasts by inflating the distributed activation differences between the conjunctions. Moreover, the stronger the response to the individual feature, the larger the consequence of divisive normalization and the resulting confounding effect. Therefore, we additionally tested whether the voxels that showed strong univariate responding could have systematically biased our multivariate tests toward significance. In a GLM, we modeled trials containing single-feature objects separately from trials containing double-feature objects. We then distributed the voxel-wise univariate response amplitudes to single-feature objects into five bins, with the first quantile representing the top 80-100% of the univariate response amplitude range, continuing until the last quantile representing the bottom 1-20% of the univariate response amplitude range. The mean t-statistic from the Trained Conjunction Contrast searchlight (as shown in Fig. 4B) was then calculated across the voxels in each bin. The results were then averaged across participants. This process was repeated for the mean t-statistic from the Novel Conjunction Contrast searchlight (as shown in Fig. 4A). Using repeatedmeasures ANOVA, we calculated a linear contrast to determine whether the mean conjunction contrast t-statistic increased or decreased systematically across the single-feature response range.

Results

1-Back Behavioral Performance

A 1-back task was used during scanning to ensure whether the participants attended the stimuli. The proportion of correctly identified repeating stimuli (hits; H) and stimuli incorrectly identified as repeats (false alarms; FA) was calculated for each session. Sensitivity (dL) was calculated from these H and FA rates by applying signal detection theory to logistic distributions: dL = ln([H(1 - FA)]/[(1 - H)FA]) (Snodgrass and Corwin 1988). Performance on the 1-back task on Session 1 (reported previously in Erez et al. 2016) was very good (dL > 7). For Session 2, overall performance was very high (dL > 7). A paired samples t-test between Session 2 Trained objects and Session 2 Novel objects revealed no significant difference between the two training conditions $(t_{(17)} = 0.19, P = 0.85, d = 0.046)$, indicating that sensitivity in the 1-back task was not higher for trained objects. However, the absence of a difference in 1-back performance between Trained and Novel objects does not speak to whether representations for these objects were unitized. Specifically, this 1-back task was designed solely to ensure that participants maintained attention, and we anticipated that performance would be universally high and thus unlikely to demonstrate differential performance across training conditions. Instead, to provide the evidence of unitization, we relied upon results from the Visual Search Task (described below) which was specifically designed for this purpose.

Visual Search Training Behavioral Performance

To ensure that some degree of unitization had occurred for the highly trained Learning Set objects, we calculated the slope of the increase in reaction time from search array set size 1 to set size 8. We predicted that untrained objects would be characterized by a steep search slope (Fig. 3A, left), where search is fast at small set sizes but slows down at large set sizes (e.g., Treisman and Sato 1990). By contrast, the unitization of trained objects is thought to allow relatively fast search even at larger set sizes, resulting in a shallow search slope (Fig. 3A, right) (Ashbridge et al. 1997). Thus, we calculated participants' visual search slopes from Day 1 to Day 6 to determine whether behavioral performance indeed reflected a decrease in visual search slope with more training (Fig. 3B). A repeated-measures ANOVA revealed a significant linear contrast in search slope from Day 1 to Day 6 ($F_{(1,17)} = 7.47$, P = 0.01, $\eta^2 = 0.31$) and no significant effects for quadratic, cubic, or any other tendency in the slope (all $F_{(1,17)} < 4.39$, all P > 0.05, all $\eta^2 < 0.21$). Direct comparison of the first and final days of training showed that the slope from Day 6 was significantly lower than the slope from Day 1 ($t_{(17)} = 2.88$, P = 0.01, d = 0.79). To investigate the role of general task practice effects, we introduced the Novel Probe Set of yellow-bodied objects on Day 6. If general practice effects were the sole cause of search improvement, these Novel Probe objects would be processed just as quickly as the Learning Set objects. However, if improvement was the result of unitized representations specific to Learning Set objects, we would find Novel Probe objects to be processed slower by comparison. Twotailed t-tests showed that on Day 6, the visual search slope for the Learning Set objects was indeed significantly lower than the search slope for Novel Probe objects ($t_{(17)} = 2.78$, P = 0.013, d = 0.37), suggesting that improvements for the Learning Set objects were likely related to unitization rather than practice. However, some effect of practice may be present as well; the visual search slope for Novel Probe objects on Day 6 was numerically lower than Learning Set objects on Day 1 ($t_{(17)} = 1.47$, P = 0.16, d = 0.41).

fMRI Analysis 1: Training-Related Changes in Conjunctive Object Representations

The first whole-brain MVPA searchlight analysis on the imaging data from fMRI Session 2 identified voxels for which there was conjunctive representation of the Session 2 Novel objects. This analysis revealed significance (voxel-level P < 0.001, uncorrected; cluster-level P < 0.001 FWE), including highly focal clusters in bilateral LOC (Fig. 4A, right peak x, y, z=26, -94, -8, respectively, z-score = 4.20, left peak x, y, z = -20, -84, -10, respectively, z-score = 4.81) and right PRC (peak x, y, z = 36, 0, -36, respectively, z-score = 3.90). For the full report of significant clusters, see Supplementary Material, Supplementary Table 1. These results replicate the findings of Erez et al. (2016). We then applied the same searchlight analysis to the Session 2 Trained objects and revealed a different spatial distribution of significant voxels. Unlike the Session 2 Novel objects, the conjunction contrast for Session 2 Trained objects identified a qualitatively more diffuse distribution of significant clusters (Fig. 4B). However, the major peaks corresponded to right FFA (peak x, y, z = -46, -8, -36, respectively, z-score = 4.47) and bilateral PPA (left peak x, y, z=14, -64, -6, respectively, z-score=4.39, left peak x, y, z = -30, -70, 4, respectively, z-score = 5.18). Finally, we conducted an MVPA searchlight to reveal voxels whose searchlight demonstrated stronger conjunctive coding for Session 2 Trained objects compared with Session 2 Novel objects (Fig. 4C). This analysis produced a focal cluster with a peak in left PRC (peak x, y, z = -38, -6, -32, respectively, z-score = 3.81), as well as a large isolated cluster that spanned bilateral posterior VVS (right peak x, y, z = -36, -52, -18, respectively, z-score = 4.61, left peak x, y, z = 36, -56, -18, respectively, z-score = 3.52) that overlapped with FFA and PPA. These data show that experience with objects recruits additional conjunctive representation in more posterior VVS regions. To investigate the functional role of the training-sensitive conjunctive code within posterior VVS, we identified the bilateral peak voxels for this Trained > Novel contrast within each participant and created 10-mm radius spheres (see ROI in the Materials and Methods section). In the next sections, this "Trained > Novel Conjunction Contrast ROI" was interrogated along with the other regions already described here.

fMRI Analysis 2: Separation of Individual Object Representations

We also asked whether the participants' representation of individual objects became more distinct once participants were more expert with each object's visual form. In the brain, increasing visual distinction would be characterized by greater separation in neural representations. Thus, we predicted that brain regions crucial for the representation of visual form would show highly separated responses to individual object exemplars after visual search training (Fig. 5B). For each ROI, we used MDS analysis to plot the individual object exemplars in a multidimensional space, with greater distances reflecting greater representational separation.

Mean interexemplar distances were calculated for each training condition in each session (Fig. 5B; e.g., mean pairwise distances between all hollow gold dots reflecting the mean interexemplar distance for Session 1 Novel objects). We first predicted no difference between the mean interexemplar distance for all Session 1 To-Be-Trained objects and the mean interexemplar distance for all Session 1 Novel objects, owing to their equivalent training status. A paired-sample t-test of the mean interexemplar distances revealed no significant difference between Session 1 To-Be-Trained objects and Session 1 Novel objects in any ROI (Fig. 5C, all $t_{(13)} < 1.56$, all P > 0.20, all d < 0.13). Interestingly, a paired-sample t-test of the mean interexemplar distances in fMRI Session 2 also revealed no significant difference between Session 2 Trained objects and Session 2 Novel objects in any ROI (all $t_{(13)} < 1.88$, P > 0.08, all d < 0.18), despite the difference in exposure and training history between the two conditions.

We then considered whether the mean interexemplar distances changed between fMRI Session 1 and fMRI Session 2 within either training condition. For each ROI, we first calculated a paired-sample t-test between the Session 1 To-Be-Trained objects and the Session 2 Trained objects, which revealed two distinct patterns of results. First, PRC showed no significant difference in mean interexemplar distance between sessions (Fig. 5C; $t_{(13)} = 1.41$, P = 0.18, d = 0.50). When considering more posterior components of the VVS-FFA, LOC, PPA, and V1-we found that interexemplar distances in fMRI Session 2 were significantly greater than those observed in fMRI Session 1 (FFA $t_{(13)} = 2.57$, P = 0.02, d = 1.02; LOC $t_{(13)} = 2.90$, P = 0.01, d = 1.19; PPA $t_{(13)} = 2.58$, P = 0.02, d = 1.04; V1 $t_{(13)} = 2.52$, P = 0.03, d = 0.98). Pairwise comparisons were also conducted on the Session 1 Novel objects and the Session 2 Novel objects. In PRC, we observed no significant difference in mean interexemplar distance between sessions ($t_{(13)} = 1.43$, P = 0.18, d = 0.55). Interestingly, the posterior components of VVS demonstrated significantly greater interexemplar distances for Session 2 Novel objects relative to Session 1 Novel objects (FFA $t_{(13)} = 2.20$, P = 0.05, d = 0.88; LOC $t_{(13)} = 2.90$, P = 0.01, d = 1.16; PPA $t_{(13)} = 2.52$, P = 0.03, d = 1.04; V1 $t_{(13)} = 2.49$, P = 0.03, d = 0.93).

In the Trained > Novel Conjunction ROI, we found the same pattern of results as we observed in other regions of posterior VVS. First, interexemplar distances did not differ by training condition within either Session 1 ($t_{(13)} = 1.85$, P=0.09, d=0.25) or Session 2 ($t_{(13)} = 1.28$, P = 0.22, d = 0.38). Second, interexemplar distances increased from fMRI Session 1 to fMRI Session 2 across both Trained ($t_{(13)} = 3.46$, P=0.004, d=0.94) and Novel objects $(t_{(13)} = 2.48, P = 0.03, d = 0.89)$. Altogether, the results show that the representations of individual objects become increasingly separated with visual training in posterior VVS. Interestingly, this object separation effect was observed in the same region that showed strong conjunctive coding for the trained objects. However, the impact of training on individual object representations persisted across the entire class of objects, including novel members of that class. Lastly, the representations of individual objects were unchanged in PRC after training, leaving open the question of whether PRC plays any role in shaping the representations in posterior VVS during learning, a question which we address in the next analysis.

fMRI Analysis 3: Changes in FC Between PRC and Posterior VVS Following Extended Learning

We predicted that learning-related changes in the distributed representations of objects would rely on input or feedback from PRC. To test this possibility, we calculated the momentto-moment connectivity between each ROI and the PRC seed region using a sliding window approach (Fig. 6A). We predicted that PRC-VVS connectivity would differ solely between Session 2 Trained objects and Session 2 Novel objects, after Visual Search Training had occurred. By contrast, we anticipated that there would be no difference in PRC-VVS connectivity between Session 1 To-Be-Trained objects and Session 1 Novel objects, owing to the fact that neither had been trained at this point.

To test these predictions, within each session, we conducted planned pairwise comparisons of connectivity with PRC between objects from each training condition. Prior to training, this analysis did not reveal any significant differences between Session 1 To-Be-Trained objects and Session 1 Novel objects within any ROI (not pictured; all $t_{(13)} < 1.60$, all P > 0.13, all d < 0.19), consistent with our predictions. Following training, however, pairwise comparisons revealed that connectivity with PRC was significantly greater during the viewing of Session 2 Novel objects relative to Session 2 Trained objects in PPA (Fig. 6B; $t_{(13)} = 2.58$, P = 0.02, d = 0.15) and V1 ($t_{(13)} = 3.80$, P = 0.002, d = 0.20), trending in LOC ($t_{(13)} = 2.01$, P = 0.07, d = 0.21), but was not significant in FFA ($t_{(13)} = 0.62$, P = 0.55, d = 0.06).

We also examined whether PRC showed significant trainingrelated changes in its interaction with the peak of conjunctive coding in posterior VVS, using the Training > Novel Conjunction ROI. As expected, in Session 1, there were no significant differences in connectivity between the To-Be-Trained objects and Novel objects ($t_{(13)} = 0.93$, P = 0.37, d = 0.11). Like several of the posterior VVS ROIs, however, this region showed significantly greater connectivity with PRC during the viewing of Session 2 Novel objects relative to Session 2 Trained objects ($t_{(13)} = 2.70$, P = 0.02, d = 0.30). In sum, PRC was more functionally connected with conjunction-sensitive posterior VVS when encountering Novel objects, pointing to a functional significance in the communication between these regions when there are new conjunctions to be learned.

Control Analysis 1: Perceptual Differences Between Objects Do Not Facilitate Training-Independent Biases in Visual Processing

An important aspect of our stimulus design was to assign one object Family with a unique feature set to each training condition, and to counterbalance those assignments across participants. Although this approach minimized the impact of any specific feature set on our present findings, the same was not done for object body color because we wanted to ensure that the stimuli in Session 2 were entirely novel and were not confused with the Trained stimulus set. However, this left the possibility that our results may arise in part from the effect of object body color per se or the physical interaction between object body color and its associated features (resulting in luminance or contrast differences, for example). Furthermore, participants were more likely to be aware of an object training status from looking at the color alone; this awareness could cue differing object viewing strategies that affected visual discrimination. Therefore, we conducted an independent control study to ensure that participants' visual discrimination performance did not differ according to body color.

First, to disentangle any practice effects for body colors that appeared later in the control experiment from performance biases arising from body color per se, we calculated a linear mixed model to determine the separate contributions of run order, color, and the interaction between run order and color on hitrate during visual search. This model revealed a significant main effect of run order on mean hitrate ($F_{(2,59,31)} = 3.82, P = 0.03$), consistent with the presence of practice effects. However, there was neither a significant main effect of color $(F_{(2,59.31)} = 1.66,$ P = 0.20) nor a significant interaction between order and color $(F_{(2,63,54)} = 0.32, P = 0.87)$. Planned pairwise comparisons based on body color also found no significant difference between the mean hitrate of any two sets of object Families (Supplementary Fig. 1A; Family 1 + 2 vs. Family 3 + 4: t₍₂₉₎ = -1.16, P = 0.26, d = 0.11; Family 3 + 4 vs. Family 5 + 6: $t_{(29)} = 0.41$, P = 0.68, d = 0.04; Family 1+2 vs. Family 5+6: $t_{(29)} = -1.65$, P = 0.11, d = 0.14; for pairwise comparisons based on run order, see Supplementary Fig. 2A).

It was necessary to determine if perceptual differences between object Families had any impact on visual search slope, given that it was our primary measure of behavioral unitization during Visual Search Training. A significant effect of color would suggest that the visual features of some Families were more readily and immediately unitized than the features of other Families, making it more difficult to interpret the source of our neuroimaging effects. Linear mixed modeling revealed a significant main effect of order $(F_{(2,58.04)} = 4.69)$, P = 0.01), consistent with the presence of practice effects. However, there was neither a significant main effect of color $(F_{(2,58.04)} = 0.53, P = 0.60)$ nor a significant interaction between order and color ($F_{(2,66.01)} = 1.12$, P=0.36). Once again, follow-up pairwise comparisons based on body color found no significant difference between the visual search slopes of any two sets of object Families (Supplementary Fig. 1B; Family 1+2 vs. Family 3 + 4: $t_{(29)} = -0.43$, P = 0.67, d = 0.05; Family 3 + 4 vs. Family 5+6: $t_{(29)} = -1.03$, P=0.31, d=0.11; Family 1+2 vs. Family 5+6: $t_{(29)} = 0.44$, P = 0.66, d = 0.06; for pairwise comparisons based on

run order, see Supplementary Fig. 2B). Together, these results suggest that perceptual differences between the object Families did not create any inherent differences in how they were visually processed. Likewise, participants' awareness of the relationship between color and object novelty was unlikely to have resulted in differing viewing strategies. Interestingly, the strategy most often reported during participant debriefing was to give some of the objects names. We speculatively interpret this naming strategy as actually being further evidence of a tendency to unitize (i.e., they have unified the features of an object under a single name, rather than considering them separately). However, this strategy applied uniquely to each object in question and did not vary systematically with our experimental conditions. In sum, the effects we observed in the main experiment most likely arose from the objects' training histories imposed by our experimental design.

Control Analysis 2: Non-linear BOLD Signal Properties Did Not Bias Multivariate Decoding

Voxel-wise Signal to Single- and Double-feature Objects

In order to test the possibility that signal saturation confounded our results, we first investigated the single- and double-feature response distributions by overlaying them with respect to each other (Supplementary Fig. 3C, top panels). This made it evident that the response distributions share a similar amplitude range, likely because the majority of the response is driven by the large object body rather than the individual features. That said that t-tests showed that the distributions did differ significantly at many points, particularly at the higher response amplitudes (Supplementary Fig. 3C, bottom panels). It was necessary to ensure that these differences were not the result of signal saturation limiting the highest double-feature response amplitudes and, therefore, negatively skewing the overall double-feature response distribution (see Supplementary Fig. 3B, blue curve, for a hypothetical example). Thus, we calculated paired Student's ttests across participants, which revealed that the skewness of the double-feature response distribution did not significantly differ from the skewness of the single-feature response distribution for either Trained objects (P=0.1) or Novel objects (P = 0.86). Indeed, any slight numerical difference in skewness we observed was positive (skewness difference for Trained objects: 0.10; Novel objects: 0.02). Importantly, we found no evidence that the double-feature response distribution was negatively skewed relative to the single-feature response distribution. This suggests that the difference in response distributions was not the result of signal saturation, but rather a more global rightward shift of the double-feature responses such as linear addition or response scaling.

The Influence of Univariate Single-Feature Response Amplitude on Multivariate Contrasts

To investigate any effects of divisive normalization, and the more general possibility where strong univariate responses may systematically bias the multivariate analyses, we tested whether the response amplitude evoked by single features was correlated with the effect strength of our conjunction contrast searchlights. Voxels were partitioned into five evenly spaced bins representing the top 80–100% of the single feature response amplitudes, continuing down to the bottom 1–20% of the single feature response amplitudes. The mean t-statistic from the multivariate conjunction contrast searchlights for Trained and Novel objects (shown in Fig. 4B and A, respectively) was calculated across the

voxels in each of these bins, and the results were averaged across participants (Supplementary Fig. 4). Using repeated-measures ANOVA, we calculated a linear contrast to determine if the searchlight t-statistics increased or decreased as a function of single-feature response amplitude. Critically, this analysis found no evidence of a positive linear relationship between univariate response amplitude and multivariate searchlight t-statistic (Trained $F_{(1,13)} = 2.15$, P = 0.17; Novel $F_{(1,13)} = 0.11$, P = 0.75). In fact, the direction of the linear relationship between the univariate response and multivariate searchlight t-statistic was numerically negative for both Trained and Novel objects. We found no significant effect for any higher order contrasts (e.g., exponential, quadratic).

Together, these observations are neither consistent with signal saturation nor divisive normalization and provide no support for the possibility of a nonlinear bias in the single-feature response confounding our multivariate effects from fMRI Analysis 1.

Discussion

It is established that certain transformations occur within the brain to allow more rapid processing of familiar experiences. In this study, we show that unitization may be a fundamental mechanism that transforms brain representations to allow more rapid processing of familiar experiences. First, we observed that conjunctive coding became stronger in posterior VVS following training. Furthermore, the representations for individual objects became more distinct from one another, an effect that was observed across both Trained and Novel objects. Finally, FC between PRC and posterior VVS changed as a function of experience such that PRC was less functionally connected with posterior VVS for highly trained objects. Below we discuss how each of these representational changes were predicted by the novel framework that we have developed.

Learning Modifies the Distributed Representation of Object Conjunctions

First, our current findings with the novel objects replicate our earlier finding that PRC and LOC possess a conjunctive code distinct from the representation of the component features themselves (Erez et al. 2016) (Fig. 4A). This is consistent with reports that damage to LOC (Behrmann and Williams 2007; Konen et al. 2011) impaired judgments requiring integration of multiple object features of similar complexity to those employed in the current study, and that the neural responses in this region tracked performance in the visual search of conjoined object dimensions (Frank et al. 2014). Moreover, amnesic patients with PRC damage were dramatically impaired when discriminating between the objects of similar complexity to those employed in the current study (e.g., Barense et al. 2005; Barense et al. 2007; Barense et al. 2012a). Critically, these deficits were observed only when it was a conjunction of complex object features that uniquely identified the solution, but not when the discrimination involved simpler stimuli-like shapes.

Having established the brain's response to novel object feature conjunctions, we turned to our critical research question: what transformations occur during extended learning to support faster discrimination of familiar objects? Similar to Czerwinski et al. (1992), our participants' behavioral patterns indicated that visual discrimination training increased visual search efficiency for trained objects relative to novel objects (Fig. 3B).

Critically, our pre and posttraining fMRI scanning also allowed us to determine whether any neural changes within VVS occurred specifically for those highly trained objects. Existing research indicates representational changes within VVS that reflect taskrelevant dimensions; for example, primate IT cortical neurons may become more selective for the conjunctions of features (Baker et al. 2002), and object representations in human ventral temporal cortex reorganize to reflect acquired contextual associations (Clarke et al. 2016). PRC is also implicated in processing stimuli with greater "lifetime" experience (Barense et al. 2010; Barense et al. 2011; Duke et al. 2016; Bowles et al. 2016). Consistent with these reports, we showed that PRC conjunctive representations were stronger for Session 2 Trained versus Session 2 Novel objects (Fig. 4C). Importantly, we further showed that entirely new clusters were recruited in posterior VVSincluding FFA and PPA-to represent Session 2 Trained feature conjunctions (Fig. 4B).

Unitization theory proposes that feature conjunctions may be "chunked" together through learning and be accessed as one unit (Goldstone 1998, 2000). While unitization has been observed in behavioral experiments across multiple visual domains, including words (e.g., Smith and Haviland 1972; Laberge 1973), faces (e.g., Valentine 1988; Tanaka and Farah 1993), novel objects (e.g., Gauthier and Tarr 1997), and even more arbitrary relational associations (e.g., Quamme et al. 2007; Parks and Yonelinas 2014), understanding of its neural underpinnings has been limited. The representationalhierarchical view posits that posterior VVS is optimized for processing simple features compared with anterior VVS/PRC, where representations are more complex and conjunctive (Bussey et al. 2002; Cowell et al. 2010). Thus, our combined understanding of these two theories predicts that unitization transforms conjunctive representations through PRC input to become more "feature-like" and represented in posterior VVS. Furthermore, past research utilizing univariate approaches has consistently highlighted the increased engagement of FFA after visual training to expert levels, or by comparing expert to naïve viewers of certain object classes (Gauthier and Tarr 1997; Gauthier et al. 1999, 2000). However, the nature of the neural computations driving that engagement had been uncertain. Our findings tie together our own model predictions with these existing observations by showing that conjunctive coding shifted to more posterior regions of the VVS, overlapping with FFA and PPA. These results are the first empirical link between unitization, the representational-hierarchical view, and their theorized neural mechanisms.

Other regions outside the VVS are thought to be important for representing the conjunctions of visual features. In particular, the role of parietal cortical subregions in the conjunctive representation of color and spatial frequency is wellsupported (Esterman et al. 2007; Baumgartner et al. 2013; Pollman et al. 2014). Interestingly, the conjunction contrast conducted for Novel objects from fMRI Analysis 1 did reveal regions outside the VVS, including clusters within the supramarginal gyrus (peak x, y, z = 50, -42, 12, respectively) and angular gyrus (peak x, y, z = -54, -54, 16, respectively; see Supplementary Table 1). These clusters overlap with other reports of subregional sensitivity to simple visual conjunctions within parietal cortex. For example, Pollman et al. (2014) reported sensitivity to color and spatial frequency fill patterns in temporoparietal junction, and Cowell et al. (2017) reported sensitivity to line contours and spatial frequency fill patterns in posterior parietal cortex. Surprisingly, the conjunction contrast yielded no significant clusters for Trained objects in these parietal subregions. This lack of training-driven conjunctive coding was also noted by Cowell et al. (2017); considering these findings together, we would speculate that parietal cortex is most involved when visual conjunctions are still novel and encoding demands are highest.

Evidence for Explicit Conjunctive Coding of Complex Object Features

The results of our experiment also help arbitrate between two predominant types of coding schemes widely thought to support conjunctive object representation. The first is a global feature-coding scheme in which individual features (e.g., A or B) are coded within visually sensitive cortical modules (Singer and Gray 1995; Eckhorn 1999). Under this scheme, conjunctive objects (e.g., AB) are coded by the co-activation of the individual modules sensitive to A and B. Thus, the whole object is quite literally the sum of its parts. On the other hand, a second coding scheme proposes that conjunctions are coded by their own cortical modules that are tuned to respond only to specific feature combinations (e.g., AB) but not to the component features themselves. This explicit conjunctive coding represents the whole object in a manner that is above and beyond the sum of its parts. Notably, finding direct evidence for each of these two representational schemes was a challenge with traditional univariate approaches. When considering a brain region's average univariate response to a conjunctive object (e.g., AB), determining how much of that response is evoked by A or B is not straightforward. However, recent studies used multivariate pattern analyses as a more sensitive tool; for example, MacEvoy and Epstein (2011) showed evidence consistent with a feature-coding scheme in LOC, where the linear combination of brain activation patterns evoked by objects A and B explained nearly all of the brain activation pattern evoked by a scene containing both objects (i.e., the whole was the sum of the parts). By contrast, Erez et al. (2016) developed a novel use of multivariate pattern analyses specifically to target the conjunctive coding scheme in the brain, and used it to show that different combinations of the same features (A + BC vs. B + AC vs. C + AB) were coded uniquely in PRC—the first direct evidence from humans for a conjunctive representation above and beyond the sum of its parts for complex objects in PRC. In the present study, we used the same approach to understand how this conjunctive code is affected by experience. To determine the likelihood that feature-coding could explain our conjunction contrast results, we directly tested whether the voxels most sensitive to single features were systematically biased to show strong effects in our multivoxel searchlights. We found no evidence to support this claim (Supplementary Fig. 4). That said, feature-coding may contribute indirectly to the conjunctive representation of complex objects and still plays an important role in visual discrimination in general. In particular, the sensitivity of a feature's multivariate pattern response to its context (e.g., A in the context of AB relative to AC) would be valuable in forming conjunctive representations of whole objects. Meanwhile, it could also have other applications. For example, sharper representation of features, in particular, contexts might improve their recognition (Moldakarimov et al. 2010). In future work, it will be important to further elucidate the functional implications of these conjunctive representations, for example, by characterizing the relationship in inter- or intra-individual differences in neural representation and

task performance, or by further elucidating the downstream mechanisms. Importantly, the representational-hierarchical view proposes that both feature-coding and conjunctive-coding schemes are present within the brain but are tuned to different levels of visual representation even for the same object (e.g., Cowell et al. 2017). In the next section, we discuss our findings in support of this prediction and highlight how the interactions between representational layers change with experience.

Extensive Training Separates the Representations of Individual Object Exemplars

We predicted that visual discrimination training would modify the brain's response not only to the conjunction of features (Fig. 4) but also to the individual objects themselves regardless of the conjunction. There is ample evidence that task-relevant dimensions shape the brain's response to visual objects, resulting in greater sensitivity to small perceptual changes (Folstein et al. 2010; Folstein et al. 2013). Beyond perception, early visual cortex increases its capacity for spatial coding and predictive coding of associated outcomes in mice and humans, respectively (Fiser et al. 2016; Hindy et al. 2016). We found that VVS responses to individual objects diverged in representational space following visual training in FFA, PPA, LOC, and V1 (Fig. 5C). Notably, this effect was also observed in a functional ROI that was highly sensitive to training-related increases in conjunctive coding (the Trained > Novel Conjunction ROI), and which we defined individually within each participant's posterior VVS. Although these findings are consistent with existing research, it might seem unusual that the visual system simultaneously embodies unitization and neural separation of these highly repeated objects. We propose that rather than being paradoxical, this property is predicted by the representational-hierarchical view.

Specifically, according to the representational-hierarchical view, conjunctions are explicitly coded—with a representation independent from the component features. The contrasting feature-coding hypothesis posits a nonlocal binding mechanism in which conjunctions are composed of individual feature representations linked by co-activation (Singer and Gray 1995; Uhlhaas et al. 2009). Thus, PRC could alternately serve as an independent substrate for explicit conjunctive representations or as a central hub in a conjunctive network that dynamically reactivates associated object features but never explicitly represents the conjunctive whole (Eckhorn 1999; Devlin and Price 2007). Conveniently, each hypothesis predicts a unique relationship between the conjunctive and individual feature "layers" of the representational hierarchy. If conjunctions are explicitly coded, modifications in the feature layer can occur without affecting the conjunctive layer. If conjunctions reflect the networks of feature representations, changes to the feature and conjunctive layers would always occur in tandem. We observed increased distinction between "individual" object representations for both Session 2 Trained and Session 2 Novel objects (Fig. 5C), whereas the "conjunctive" coding was stronger for Session 2 Trained relative to Session 2 Novel objects (Fig. 4C). This suggests that posttraining modifications can occur independently in different layers of the representational hierarchy, and argues against the alternate view that featurecoding or feature sharpening can explain the observations of a conjunctive response. Convergent work exploring the transition from feature-coding to conjunction-coding in the VVS showed evidence for a hierarchical coding scheme that was not only representationally separate but also anatomically separate along an anterior-posterior axis (Cowell et al. 2017). Our complementary findings demonstrate that the neural architecture supporting feature-level coding can also be reorganized to support the coding of unitized conjunctions through experience.

Why did feature separation occur even for Session 2 Novel objects, when one might have predicted such changes only for the Session 2 Trained objects? The representational-hierarchical framework focuses on how exposure shapes object representations and is agnostic to how these effects might generalize to new material that does not directly overlap in terms of either the features or their conjunctions. However, there are several post hoc explanations for why generalization is a realistic possibility in the present study. First, the demands of the 1-back task behoove participants to distinguish between objects regardless of past training (e.g., object A and object AB should never be confused despite a shared feature). Second, participants may be acquiring expertise in this class of objects as a whole, which may generalize to novel members of that class. Behavioral studies broadly corroborate this assertion (e.g., Tanaka et al. 2005; Wong et al. 2009; Bukach et al. 2012). In support of this view, the visual search slopes of Novel Probe (yellow) objects on Day 6 were numerically shallower than the search slopes of the Learning Set (blue) objects on Day 1 (d = 0.41; Fig. 3B), suggesting that there was some training-related generalization to the Novel Probe objects. Moreover, our stimuli were likely conducive to transfer of expertise; despite being different colors, Session 2 Trained objects and Session 2 Novel objects shared the same body shape, and their features occupied similar regions of visual space.

On the other hand, it is possible that Trained objects could have exhibited greater learning-related neural separation relative to Novel objects, but we lacked sufficient control over which dimensions each participant found to be most salient or could not guarantee that those dimensions were equally salient to start. Moreover, differences in perceived saliency across the unique Novel objects from fMRI Sessions 1 and 2 could have masked any conditional effect of training in a manner subjective to each participant. It would be possible to be certain of the specific contributions to feature separation in a separate experiment where the diagnostic features within an object Family were better controlled, or if we ensured that the Novel objects were actually the same objects across fMRI sessions but did not undergo visual training. For example, an experiment based on the present study could employ a modified version of the Visual Search Task where the list of possible distractors were restricted for each object Family, such that some features would never or rarely be foiled during training. In this way, one could manipulate the level of training of specific features while controlling the amount of visual exposure and keeping all conjunctions intact. A per-Family separation score would capture the neural distance between more trained versus less trained features within the same Family of objects. This additional level of granularity may be necessary to detect the unique effect of Visual Search Training within each object Family and individual participant.

Learning is Supported by Interactions Between PRC and Posterior VVS

Finally, we asked whether different levels of the VVS hierarchy become more functionally connected to support experiencerelated representational changes. As a starting point, we considered PRC due to its persistent role in the conjunctive representation both before and after visual training, and established

anatomical and functional evidence for a connection between PRC and posterior VVS (Suzuki and Amaral 1994; Clavagnier et al. 2004; Peterson et al. 2012; Cacciamani et al. 2017). We predicted that PRC could either mediate the formation of unitized conjunctive representations in posterior VVS during learning, or facilitate the retrieval of unitized conjunctive representations after learning. In the former case, PRC would be more functionally connected to posterior VVS during novel objects (in order to take and unitize information about individual features); in the latter case, connectivity would be greater during trained objects (in order to facilitate rapid retrieval of single features). Consistent with the former prediction, we found that FC between PRC and V1 and between PRC and PPA was highest during Novel objects. Importantly, the Trained > Novel Conjunction Contrast ROI—representing the peak of training-related increases in conjunctive coding—also showed stronger FC with PRC during Novel objects. Critically, we observed none of these effects in fMRI Session 1, before any Visual Training had occurred. Together, these are strong evidence for a role of novelty in gating PRC-VVS interactions that underlie unitization.

Notably, computational models of the representational hierarchy have suggested that conjunctive representations can form over many trials of input without any interaction between the PRC and posterior layers (Cowell et al. 2006; Sadil and Cowell 2016), and one amnesic case with PRC damage was able to spontaneously unitize relational stimuli (Ryan et al. 2013). It is an empirical question whether unitized conjunctions in posterior VVS formed with PRC input are fundamentally different from those formed without. Furthermore, damage to PRC may limit the speed and complexity of conjunctive representations to be formed while preserving the formation of less complex conjunctions.

Conclusions

Although it is commonly accepted that our ability to perceive complex objects improves with repeated experience, the neural transformations in support of this behavior are relatively unknown. This study tested a functional model that unifies two established theories: the representational-hierarchical view and the theory of unitization. Importantly, this unified model makes unique predictions about the effect of experience on the spatial distribution of conjunctive representations across the VVS and the communication between VVS regions—predictions that could not have been generated by either theory alone. Using fMRI and multivariate analysis techniques, we have provided the first compelling data demonstrating how the visual system elevates feature conjunctions according to task demands and distributes their representation to ensure their rapid retrieval in the future.

Notes

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. We also thank Thaddea Chua, Sarah J. Wong, and Layan Elfaki for their data collection efforts.

Supplementary Material

Supplementary material can be found at Cerebral Cortex online.

References

- Ashbridge E, Walsh V, Cowey A. 1997. Temporal aspects of visual search studied by transcranial magnetic stimulation. *Neuropsychologia*. 35(8):1121–1131.
- Baker CI, Behrmann M, Olson CR. 2002. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. Nat Neurosci. 5(11):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. *Journal of Neuroscience*. 25(44):10239–10246.
- Barense MD, Gaffan D, Graham KS. 2007. The human medial temporal lobe processes online representations of complex objects. *Neuropsychologia*. 45(13):2963–2974.
- 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(3):389–401.
- 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(10):3052–3067.
- Barense MD, Groen IIA, Lee ACH, Yeung LK, Brady SM, Gregori M, Kapur N, Bussey TJ, Saksida LM, Henson RN. 2012a. Intact memory for irrelevant information impairs perception in amnesia. Neuron. 75(1):157–167.
- Barense MD, Ngo JKW, Hung LHT, Peterson MA. 2012b. Interactions of memory and perception in amnesia: the figureground perspective. Cereb Cortex. 22(11):2680–2691.
- 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(12):821–32.
- Baumgartner F, Hanke M, Geringswald F, Zinke W, Speck O, Pollman S. 2013. Evidence for feature binding in the superior parietal lobule. *Neuroimage*. 68:173–180.
- Behrmann M, Marotta J, Gauthier I, Tarr MJ, McKeeff TJ. 2005. Behavioral change and its neural correlates in visual agnosia after expertise training. J Cogn Neurosci. 17(4): 554–568.
- Behrmann M, Williams P. 2007. Impairments in pair-whole representations of objects in two cases of integrative visual agnosia. *Cogn Neuropsychol.* 24(7):701–730.
- Bowles B, Duke D, Rosenbaum SR, McRae K, Kohler S. 2016. Impaired assessment of cumulative lifetime familiarity for object concepts after left anterior temporal-lobe resection that includes perirhinal cortex but spares the hippocampus. *Neuropsychologia*. 90:170–179.
- Bukach CM, Gauthier I, Tarr MJ, Kadlec H, Barth S, Ryan E, Turpin J, Bub DN. 2012. Does acquisition of Greeble expertise in prosopagnosia rule out a domain-general deficit. Neuropsychologia. 50(2):289–304.
- Bussey TJ, Saksida LM, Murray EA. 2002. Perirhinal cortex resolves feature ambiguity in complex visual discriminations. Eur J Neurosci. 15(2):365–374.
- Cacciamani L, Wager E, Peterson MA, Scalf PE. 2017. Age-related changes in perirhinal cortex sensitivity to configuration and part familiarity and connectivity to visual cortex. Front Aging Neurosci. 9:291.
- Clarke A, Pell PJ, Ranganath C, Tyler LK. 2016. Learning warps object representations in the ventral temporal cortex. J Cogn Neurosci. 28(7):1010–1023.

- Clavagnier S, Falchier A, Kennedy H. 2004. Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. *Cogn Affect Behav Neurosci.* 4(2):117–126.
- Cowell RA, Bussey TJ, Saksida LM. 2006. Why does brain damage impair memory? A connectionist model of object recognition memory in perirhinal cortex. J Neurosci. 26(47):12186–12197.
- Cowell RA, Bussey TJ, Saksida LM. 2010. Components of recognition memory: dissociable cognitive processes or just differences in representational complexity? *Hippocampus*. 20(11):1245–1262.
- Cowell RA, Leger KR, Serences JT. 2017. Feature-coding transitions to conjunction-coding with progression through human visual cortex. J Neurophysiol. 118(6):3194–3214.
- Cusack R, Vicente-Grabovetsky A, Mitchell DJ, Wild CJ, Auer T, Linke AC, Peelle JE. 2015. Automatic analysis (AA): efficient neuroimaging workflows and parallel processing using Matlab and XML. Front Neuroinform. 8:90.
- Czerwinski M, Lightfoot N, Shiffrin RM. 1992. Automatization and training in visual search. *Am J Psychol*. 105(2):271–315.
- Dale AM. 1999. Optimal experimental design for event-related fMRI. Hum Brain Mapp. 8:109–114.
- 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, pp. 267–299.
- Devlin JT, Price CJ. 2007. Perirhinal contributions to human visual perception. Curr Biol. 4(17):1484–1488.
- Duke D, Martin CB, Bowles B, McRae K, Kohler S. 2016. Perirhinal cortex tracks degree of recent as well as cumulative lifetime experience with object concepts. *Cortex*. 89:61–70.
- Eckhorn R. 1999. Neural mechanisms of visual feature binding investigated with microelectrodes and models. Vis Cogn. 6(3– 4):231–265.
- Erez J, Cusack R, Kendal W, Barense MD. 2016. Conjunctive coding of complex object features. Cereb Cortex. 26(5):2271–2282.
- Esterman M, Verstynen T, Robertson LC. 2007. Attenuating illusory binding with TMS of the right parietal cortex. *Neuroimage*. 35(3):1247–1255.
- Fiser A, Mahringer D, Oyibo HK, Petersen AV, Leinweber M, Keller GB. 2016. Experience-dependent spatial expectations in mouse visual cortex. Nat Neurosci. 19:1658–1664.
- Folstein JR, Gauthier I, Palmeri TJ. 2010. Mere exposure alters category learning of novel objects. Front Psychol. 1(40):1–6. doi: 10.3389/fpsyg.2010.00040.
- Folstein JR, Gauthier I, Palmeri TJ. 2013. Category learning increases discriminability of relevant object dimensions in visual cortex. *Cereb Cortex*. 23:814–823.
- 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(4):1201–1211.
- Gauthier I, Tarr MJ. 1997. Becoming a "Greeble" expert: exploring mechanisms for face recognition. Vis Res. 37(12):1673–1682.
- Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC. 1999. Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. Nat Neurosci. 2(6):568–573.
- Gauthier I, Skudlarski P, Gore JC, Anderson AW. 2000. Expertise for cars and birds recruits brain areas involved in face recognition. Nat Neurosci. 3(2):191–197.
- Goldstone RL. 1998. Perceptual learning. Annu Rev Psychol. 49:585–612.

- Goldstone RL. 2000. Unitization during category learning. J Exp Psychol Hum Percept Perform. 26(1):86–112.
- Gross CG. 1992. Representation of visual stimuli in inferior temporal cortex. Philos Trans Royal Soc London B Biol Sci. 335(1273):3–10.
- Hinds OP, Rajendran N, Polimeni JR, Augustinack JC, Wiggins G, Wald LL, Diana Rosas H, Potthast A, Schwartz EL, Fischl B. 2008. Accurate prediction of V1 location from cortical folds in a surface coordinate system. *Neuroimage*. 39(4): 1585–1599.
- Heeger DJ. 1992. Normalization of cell responses in cat striate cortex. Vis Neurosci. 9:181–197.
- Hindy NC, Ng FY, Turk-Browne NB. 2016. Linking pattern completion in the human hippocampus to predictive coding in visual cortex. Nat Neurosci. 19(5):665–667.
- Hutchison RM, Womelsdorf T, Allen EA, Bandettini PA, Calhoun VD, Corbetta M, Penna SD, Duyn JH, Glover GH, Chang C. 2013. Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*. 80:360–378. doi: 10.1016/j.neuroimage.2013.05.079.
- Kay KN, Rokem A, Winawer J, Dougherty RF, Wandell BA. 2013. GLMdenoise: a fast, automated technique for denoising taskbased fMRI data. Front Neurosci. 7(247):1–15.
- Konen CS, Behrmann M, Nishimura M, Kastner S. 2011. The functional neuroanatomy of object agnosia: a case study. *Neuron*. 71(1):49–60.
- Kriegeskorte N, Goebel R, Bandettini P. 2006. Informationbased functional brain mapping. Proc Natl Acad Sci USA. 103(10):3863–3868.
- Kriegeskorte N, Mur M, Bandettini P. 2008. Representational similarity analysis—connecting the branches of systems neuroscience. Front Syst Neurosci. 2(4):1–28.
- Laberge D. 1973. Attention and the measurement of perceptual learning. Cognition. 1(3):268–276.
- Lee ACH, Rudebeck SR. 2010. Human medial temporal lobe damage can disrupt the perception of single objects. J Neurosci. 30(19):6588–6594.
- Linke AC, Vicente-Grabovetsky A, Cusack R. 2011. Stimulus-specific suppression preserves information in auditory short-term memory. Proc Nat Acad Sci USA. 108(31):12961–12966.
- MacEvoy SP, Epstein RA. 2009. Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr Biol.* 19:1–5.
- MacEvoy SP, Epstein RA. 2011. Constructing scenes from objects in human occipitotemporal cortex. Nat Neurosci. 14:1323–1329.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Tootell RB. 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Nat Acad Sci USA. 92(18): 8135–8139.
- Moldakarimov S, Bazhenov M, Sejnowski TJ. 2010. Representational sharpening can explain perceptual priming. *Neural Comput.* 22(5):1312–1332.

- Parks CM, Yonelinas AP. 2014. The importance of unitization for familiarity-based learning. J Exp Psychol Learn Mem Cogn. 41(3):881.
- 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(10):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.
- Quamme JR, Yonelinas AP, Norman KA. 2007. Effect of unitization on associative recognition in amnesia. *Hippocampus*. 17(3):192.
- Reddy L, Kanwisher N. 2007. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Curr Biol.* 17(23):2067–2072.
- Riesenhuber M, Poggio T. 1999. Hierarchical models of object recognition in cortex. Nat Neurosci. 2(11):1019–1025.
- Ryan JD, Moses SN, Barense MD, Rosenbaum RS. 2013. Intact learning of new relations in amnesia as achieved through unitization. J Neurosci. 33(23):9601–9613.
- Sadil PS, Cowell RA. 2016. A computational model of perceptual and mnemonic deficits in medial temporal lobe amnesia. J Cogn Neurosci. 29(6):1075–1088.
- Singer W, Gray CM. 1995. Visual feature integration and the temporal correlation hypothesis. Annu Rev Neurosci. 18:555–586.
- Smith EE, Haviland SE. 1972. Why words are perceived more accurately than nonwords: inference versus unitization. J Exp Psychol. 92(1):59–64.
- Snodgrass JG, Corwin J. 1988. Pragmatics of measuring recognition memory: applications to dementia and amnesia. J Exp Psychol Gen. 117(1):34–50.
- Suzuki WA, Amaral DG. 1994. Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. J Comp Neurol. 350(4):497–533.
- Tanaka JN, Farah MJ. 1993. Parts and wholes in face recognition. Quar JExp Psychol A Hum Exp Psychol. 46(2):225–245.
- Tanaka K. 1996. Inferotemporal cortex and object vision. Annu Rev Neurosci. 19:109–139.
- Tanaka JW, Curran T, Sheinberg DL. 2005. The training and transfer of real-world perceptual expertise. Psychol Sci. 16(2):145–151.
- Treisman A, Sato S. 1990. Conjunction search revisited. J Exp Psychol Hum Percept Perform. 16(3):459–478.
- Uhlhaas PJ, Pipa G, Lima B, Melloni L, Neuenschwander S, Nikolic D, Lin S. 2009. Neural sychrony in cortical networks: history, concept and current status. Front Integ Neurosci. 3(17):1–19.
- Valentine T. 1988. Upside-down faces: a review of the effect of inversion upon face recognition. Br J Psychol. 79(4):471–491.
- Watson HC, Wilding EL, Graham KS. 2012. A role for perirhinal cortex in memory for novel object-context associations. J Neurosci. 32(13):4473–4481.
- Wong AC, Palmeri TJ, Gauthier I. 2009. Conditions for facelike expertise with objects: becoming a Ziggerin expert—but which type? Psychol Sci. 20(9):1108–1117.