NeuroImage 134 (2016) 170–179

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Typicality sharpens category representations in object-selective cortex

Marius Cătălin Iordan [a,*], Michelle R. Greene [a], Diane M. Beck [b], Li Fei-Fei [a]

[a] Department of Computer Science, Stanford University, Stanford, CA 94305, USA
[b] Beckman Institute and Department of Psychology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

ARTICLE INFO

Article history:
Received 20 January 2016
Revised 12 March 2016
Accepted 5 April 2016
Available online 12 April 2016

Keywords:
Categorization
fMRI
Object
Typicality

ABSTRACT

The purpose of categorization is to identify generalizable classes of objects whose members can be treated equivalently. Within a category, however, some exemplars are more representative of that concept than others. Despite long-standing behavioral effects, little is known about how typicality influences the neural representation of real-world objects from the same category. Using fMRI, we showed participants 64 subordinate object categories (exemplars) grouped into 8 basic categories. Typicality for each exemplar was assessed behaviorally and we used several multi-voxel pattern analyses to characterize how typicality affects the pattern of responses elicited in early visual and object-selective areas: V1, V2, V3v, hV4, LOC. We found that in LOC, but not in early areas, typical exemplars elicited activity more similar to the central category tendency and created sharper category boundaries than less typical exemplars, suggesting that typicality enhances within-category similarity and between-category dissimilarity. Additionally, we uncovered a brain region (cIPL) where category boundaries favor less typical categories. Our results suggest that typicality may constitute a previously unexplored principle of organization for intra-category neural structure and, furthermore, that this representation is not directly reflected in image features describing natural input, but rather built by the visual system at an intermediate processing stage.

1. Introduction

The purpose of categorization is to identify generalizable classes of objects whose members can be treated equivalently. Within a category, however, some exemplars are more representative of that concept than other members of the same category. This typicality effect usually manifests behaviorally as increased speed of recognition, as well as lower error rates for verifying category membership of the more typical item (Posner and Keele, 1968; Rosch, 1973; Rosch and Mervis, 1975). Despite well-studied behavioral effects, little is known about how typicality influences the neural representation of objects from the same category: for example, why are some dog exemplars more representative of the category "dog" than others and where can we find evidence for this distinction in the brain?

Previous investigations of the neural basis for typicality have employed category learning paradigms over artificially constructed categories (Aizenstein et al., 2000; Zeithamova et al., 2008; Davis et al., 2012a,b; Davis and Poldrack, 2014). By contrast, our environment contains tens of thousands of distinct object categories (Biederman, 1987; Deng et al., 2009). Furthermore, considerable evidence suggests that perceived typicality is reflected in how fast and how accurately we perceive many such real-world objects and categories (Posner and Keele, 1968; Rosch, 1973; Rosch and Mervis, 1975). Thus, the overarching goal of our present work is to investigate how the typicality of real-world object categories affects their representation in human visual cortex.

Many theories and cognitive models have been proposed for the instantiation of typicality as a dimension of object representation in human categorization (for reviews, see e.g. Ashby and Maddox, 1993, 2005; Minda and Smith, 2002; Abbott et al., 2012), however, a clear neural correlate of these models has yet to be identified. Nevertheless, in virtually all such models, distinct objects are defined as points in a multidimensional psychological space and similarity (in terms of features or properties) between such items belonging to the same or different categories represents the defining characteristic by which typicality (and categorization itself) is instantiated. In the spirit of this observation, we set out to test one of the earliest and most fundamental hypotheses regarding the instantiation of typicality relationships between exemplars in a given category: the family resemblance hypothesis first put forward by Rosch and Mervis (1975). Their proposed model states that highly typical members of a category are those that share most features in common with other members of that category (i.e. a typical subordinate level exemplar, such as a Golden Retriever, is highly representative of the basic level category 'dog'), while simultaneously sharing the fewest features in common with other categories in a similar

* Corresponding author at: Computer Science Department, Stanford University, 353 Serra Mall, Rm. 240, Stanford, CA 94305, USA.
E-mail addresses: mci@cs.stanford.edu (M.C. Iordan), mrgreene@stanford.edu (M.R. Greene), dmbeck@illinois.edu (D.M. Beck), feifeili@cs.stanford.edu (L. Fei-Fei).

http://dx.doi.org/10.1016/j.neuroimage.2016.04.012
1053-8119/

semantic space (i.e. with other basic level categories within the same superordinate category; e.g. Golden Retrievers would share very few features in common with cats).

Investigating hypotheses such as this one is challenging in the real-world domain mainly because the sheer number of categories in our environment is estimated to be in the tens of thousands (Biederman, 1987) and because controlling for the features of natural visual stimuli is notoriously difficult. In our present experiment, we put forward the first attempt to push beyond small-scale, artificial, hand-designed datasets for investigating how typicality modulates neural representations by leveraging a large-scale taxonomically structured image database (ImageNet, Deng et al., 2009), along with employing a method for obtaining high-throughput behavioral rankings (the Amazon Mechanical Turk platform). As such, we are now able to test directly whether brain regions exist where the family resemblance hypothesis represents a guiding principle for the neural intra-class organization of a large set of real-world object categories and, furthermore, compare this organization against the corresponding low-level visual feature representation of the over one thousand images we used as stimuli in our study.

To this end, we performed a passive viewing fMRI experiment in which participants viewed color photographs from 64 subordinate level object categories grouped into 8 basic level categories. The typicality of each subordinate category (hitherto referred to as an "exemplar") within its corresponding basic category (hitherto referred to as a "category") was ranked behaviorally. The family resemblance model was originally defined using a semantic feature space: e.g. the category 'dog' is exemplified by features such as 'has-tail', 'wags-tail', and 'is-furry'; and an exemplar which possesses more of these features would be rated as more typical. Although Rosch's family resemblance hypothesis has been well received, it has been difficult to find definitive evidence for it primarily because the feature space used by the brain is unknown. Here, we set out to investigate this question in the domain of neural activation patterns, where we can remain agnostic as to the nature of the feature spaces, semantic or otherwise, in which object categories are represented. Multi-voxel pattern analyses allow us to characterize the similarity between neural patterns elicited by these categories throughout human visual cortex, without making any explicit assumptions regarding the building blocks of the feature spaces themselves. As such, we found that in object-selective regions of occipito-temporal cortex, but not in early visual areas, typical exemplars were more similar to the central tendency of the category and created significantly sharper category boundaries than less typical exemplars, suggesting that typicality enhances category cohesion (within-category similarity) and category distinctiveness (between-category dissimilarity). Thus, we present the first evidence that typicality modulates neural representations of real-world object categories in object-selective cortex in a manner consistent with the family resemblance hypothesis. Interestingly, using a whole-brain analysis, we also uncovered the first evidence of a brain region where category boundaries favor less typical categories (cIPL). Taken together, these findings suggest that the two extremes of the behavioral typicality continuum may simultaneously exert separate influence on the neural representation of real-world object categories across human visual cortex, and moreover, that typicality may constitute a previously unexplored principle of organization for intra-category neural structure, one that is likely built by the visual system at an intermediate processing stage, rather than inherited from low-level features of our input.

2. Materials and methods

2.1. Constructing a behaviorally-normed category set

The goal of our experiment was to test the family resemblance hypothesis (Rosch and Mervis, 1975) which posits that highly typical members of a category share the most features in common with other members of that category, while simultaneously sharing the fewest features in common with members of semantically related categories. To test this model appropriately, we required a set of basic level categories (e.g. dog, car), each comprising multiple subordinate level categories (exemplars, e.g. *Chihuahua*, sedan) for which perceived typicality could be assessed behaviorally.

In our experiment, we started with a four-tiered taxonomic hierarchy comprising the following putative levels: two domain level categories (natural, man-made), four superordinate level categories (animals, plants, musical instruments, vehicles), sixteen basic level categories (e.g. bird, cat, dog, fish for 'animals'), and one hundred and twenty-eight subordinate level categories (e.g. *Chihuahua*, stealth plane, parsley). Subsequently, we assessed the entry levels in each of our four superordinate tiers. We performed a match-to-category behavioral experiment in which we asked participants to verify whether each image belonged to its subordinate, basic, superordinate, or domain level category. We found that, of our four putative superordinate categories, 'animals' and 'vehicles' were the only ones who adhered strongly to the putative hierarchy, whereas plants and musical instruments varied across disparate taxonomic tiers and, for some of their categories, the basic level was situated either at a more general or more specific tier than their putative designation (e.g. putative basic levels 'wind instruments', 'string instruments', 'garden plants' closer to superordinate level; putative superordinate level 'plants' closer to basic level; putative superordinate 'musical instruments' closer to domain level; see Supplementary material and Supplementary Figs. S1 & S2). Therefore, to maintain a consistent, verified hierarchy, we selected a subset of our original dataset comprising eight basic level categories (dogs, cats, birds, fish; cars, boats, planes, trains) and sixty-four subordinates (eight for each basic category, e.g. *Chihuahua*, stealth plane, etc.). This hierarchy has the added advantage that it contains equal numbers of natural/animate and man-made/inanimate categories, a distinction known to affect representations of object categories in human visual cortex (Connolly et al., 2012; Konkle and Caramazza, 2013).

Subsequently, we used ImageNet (Deng et al., 2009) to collect 16 distinct images containing objects of interest from each of our sixty-four subordinate level categories; i.e. if the subordinate category is pugs, then we showed 16 distinct photographs of pugs. Pictures were cropped to feature the objects prominently and centrally within a square region (400×400 pixels in size) and included their natural background. Within each subordinate category, the images varied greatly in color and pose. Representative images from each of our 64 categories are shown in Fig. 1.

2.2. Behavioral experiment: typicality rankings

2.2.1. Participants and materials

40 participants were recruited on Amazon's Mechanical Turk platform (AMT) from a pool of trusted US-based participants with at least 2000 previously accepted AMT results at a minimum of 98% approval. Participants completed the study from their own personal computing device.

2.2.2. Experimental procedures

Each of the AMT hits contained 28 trials comprising each possible pairwise comparison between the eight subordinate categories within a particular basic category. In each trial, participants viewed a randomly drawn image from two subordinate categories and were asked to indicate by clicking which image was the most typical of its corresponding basic category. Ten individual participants ranked each basic category, with each participant ranking a median of six basic level categories overall. Participants were compensated $0.50 per hit and each hit took an average of 88 s to complete.

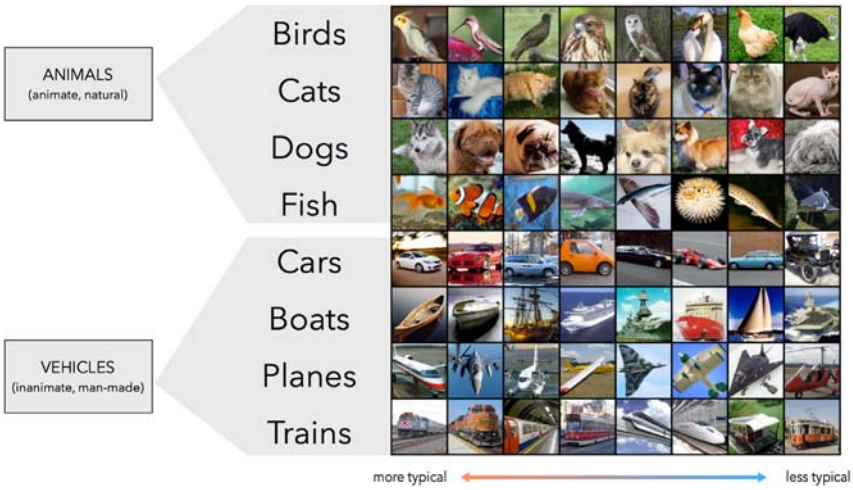

Fig. 1. Typicality ranked stimulus set. Our stimulus set comprised 8 subordinate level exemplars from each of 8 basic level categories. Participants were shown 16 images from each exemplar, varying in pose and color (only one representative image is shown above). Within each basic category, exemplars are organized according to behavioral typicality from the most typical (left) to the least typical (right): e.g. airliners (rank 1) and fighter planes (rank 2) were judged to be much more typical examples of planes than stealth planes (rank 7) and gyrocopters (rank 8).

2.2.3. Data analysis

Pairwise typicality rankings for the eight subordinates in each basic category were obtained. We computed the percentage of times each subordinate was chosen as the more typical item in a pair and used this quantity to order subordinates according to their typicality in each basic category independently. We also recorded a high value for the inter-subject reliability of the collected typicality rankings (75% ± 2%, mean ± s.e.m.; see Supplementary Fig. S3).

2.3. fMRI experiment

2.3.1. Participants

12 volunteers (2 females, ages 24–32, including authors M.C.I. and M.R.G.) with no past history of psychiatric or neurological disorders and normal or corrected-to-normal vision participated in this experiment. Participants gave informed written consent in compliance with procedures approved by the Stanford University Institutional Review Board. Except for the participating authors, all subjects received financial compensation. One participant was subsequently rejected from our analyses due to our inability to satisfactorily identify their regions of interest using the localizer scanning procedures detailed in the corresponding section below.

2.3.2. Scanning parameters and preprocessing

Imaging data were acquired with a 3 Tesla G.E. Healthcare scanner. A gradient echo, echo-planar sequence was used to obtain functional images (volume repetition time (TR), 2 s; echo time (TE), 30 ms; flip angle, 80°; matrix, 128 × 128 voxels; FOV, 20 cm; 29 oblique 3 mm slices with 1 mm gap; in-plane resolution, 1.56 × 1.56 mm). We also collected a high-resolution (1 × 1 × 1 mm voxels) structural scan (SPGR; TR, 5.9 ms; TE, 2.0 ms; flip angle, 11°) in each scanning session. The functional data were spatially aligned to compensate for motion during acquisition and each voxel's intensity was converted to percent signal change relative to the temporal mean of that voxel using the AFNI software package (http://afni.nimh.nih.gov/afni). To perform our analyses, we computed the average voxel activity for each block. We did not perform any smoothing.

2.3.3. Experimental procedure

Images were presented centrally subtending 21°×21° visual angle and were superimposed on an equiluminant gray background. We used a back-projection system (Optoma Corporation) operating at a resolution of 1024 × 768 pixels at 75 Hz. Participants performed 2 sessions, 8 runs each, with 16 blocks per run and 8 images per block. Each block consisted of a 500 ms fixation cross presented centrally, followed by 8 consecutive stimulus presentations from the same subordinate level category, with a 12 s gap between the blocks. Each image was presented for 160 ms, followed by a 590 ms blank gray screen. Subjects were asked to maintain fixation at the center of the screen, and respond via button-press whenever an image was repeated (one-back task, 0–2 repetitions per block). Over the course of the experiment, each participant viewed 2 blocks from each of the subordinate level categories. The order of blocks, the number of repetitions in each block, and the images in each block were counter-balanced across runs and between subjects. The experiment was implemented in MATLAB (www.mathworks.com), using the Psychophysics toolbox extension (Brainard, 1997; Pelli, 1997).

2.3.4. Regions of interest (ROIs)

The positions and extents of each participant's lateral occipital complex (LOC) were obtained using standard localizer runs conducted in a separate fMRI session. Participants completed two runs, each with 12 blocks drawn equally from six categories: child faces, adult faces, indoor scenes, outdoor scenes, objects (abstract sculptures with no semantic meaning), and phase-scrambled objects. Blocks were separated by 12 s fixation cross periods and comprised 12 image presentations, each of which consisted of images presented for 900 ms, followed by a 100 ms fixation cross. Each image was presented exactly once, with the exception of two images during each block that were repeated twice in a row. Subjects were asked to maintain fixation at the center of the screen and respond via button press whenever an image was repeated. To avoid any issues related to intrinsic variability in signal reliability across our participant pool, we selected fixed-volume ROIs across all our participants. The volume of LOC in mm^3 was chosen conservatively, based on sizes previously reported in the literature, accounting for resolution differences between studies (Golarai et al., 2007; Walther et al., 2009; Iordan et al., 2015). Accordingly, LOC was defined as the top 500 voxels bilaterally near the inferior occipital gyrus that responded to an objects > scrambled objects GLM contrast.

To determine the locations of early visual areas V1, V2, V3v, and hV4, we used a standard retinotopic mapping protocol in a separate experiment, in which a checkerboard pattern undergoing contrast reversals at 5 Hz moved through the visual field in discrete increments (Sayres and Grill-Spector, 2008). First, a wedge subtending an angle of 45° from fixation was presented at 16 different polar angles for 2.4 s each.

Next, an annulus subtending 3° of visual angle was presented at 15 different radii for 2.4 s each. Each subject passively observed two runs of 6 cycles in each condition, yielding 512 timepoints per subject. The locations and extents of early visual areas were delineated on a flattened cortical surface for each subject, using a horizontal vs. vertical meridian general linear test, which gave the boundaries between retinotopic maps.

We aligned the positions of the ROIs to the experimental sessions using the AFNI software package (http://afni.nimh.nih.gov/afni), by first aligning the structural scans between sessions with sub-millimeter precision, and then applying the alignment transformation to the ROI positions. Percent signal change was then extracted for each voxel in each ROI and these vectors were submitted to the similarity analyses described next.

2.4. fMRI data analysis

2.4.1. Correlation advantage

First, we assessed whether the most or the least typical exemplars in each category were more similar to the central category tendency. To this end, for each basic category, we used the average neural patterns of all exemplars as a proxy for the central category tendency representation. This definition is similar to that of a putative prototype for that category (Sigala and Logothetis, 2002). We then computed the correlation (Pearson's r) between this category central tendency, on the one hand, and the most and least typical subordinates in each basic category, on the other hand. We hypothesized that if the family resemblance hypothesis is upheld, then the most typical subordinate will be more similar (correlated in its elicited pattern of activation) to the central category tendency than the least typical subordinate. Additionally, we computed a version of this analysis where we omitted from the computation of the central tendency the most typical and least typical exemplars (leaving only the six middle-typicality exemplars in each category). Results were similar, regardless of the method used to compute the central category tendency. Throughout our analyses, we chose to focus on Pearson correlation as a straightforward, scale-invariant measure of similarity of neural patterns, which has the ability to normalize across differences in mean activation level between stimuli and is therefore less susceptible to such variation across a large set of object categories.

2.4.2. Category boundary effect

Next, we assessed whether typical exemplars share fewer features in common with other categories than less typical exemplars. Here, we refer to neural features (as measured by voxel activity levels) and we make no assumption that the features are semantic or otherwise (Clarke and Tyler, 2014), only that multi-voxel patterns reflect some underlying feature space. By measuring similarity of brain activity patterns we aim to bridge the gap between the two types of features, positing that similarity in one descriptive space (voxels) is a good proxy for similarity in the other (internal feature representation). We hypothesized that if this is the case, then categories defined solely by relatively higher typicality exemplars would be more distinguishable from one another than categories comprising only less typical exemplars. To this end, for each ROI and each subject, we split our dataset into two halves comprising the four most typical and four least typical exemplars, respectively, from each category. We then computed a category boundary effect measure separately for each of the two halves of our dataset. We defined the category boundary effect identically to previous work (Kriegeskorte et al., 2008; Iordan et al., 2015) as the difference between within-category similarity and between-category similarity, averaged across all categories considered. For each basic level category, we computed within-category similarity as the average correlation (Pearson's r) between neural patterns elicited by within-category pairs of blocks (e.g. for 'dogs', this quantity is defined as the average correlation between voxel activations for any two blocks where any type of dog was shown). Similarly, we computed between-category similarity as the average correlation between neural patterns elicited by between-category pairs of blocks across basic level categories (e.g. for 'dogs', this quantity is defined as average correlation between voxel activations for a block where dogs were shown and another block where, for example, planes were shown). We performed each of these analyses for each subject and ROI separately. We used this measure to quantify how well categories are separated in the neural space of representation, given their behavioral typicality.

2.4.3. Low-level feature analysis

To show that the effects in the correlation advantage and category boundary effect analyses above, are not solely due to low-level image features, we also performed analogous computations for image descriptor features extracted from our stimulus images: LAB color histograms, GIST (Oliva and Torralba, 2001), and multi-scale Gabor wavelet features (Kay et al., 2008). Color histograms were represented using LAB color space. For each image, we created a two-dimensional histogram of the a* and b* channels using 64 bins per channel. We then averaged these histograms over each of the 16 distinct stimuli in each subordinate category, such that each subordinate was represented as a 4096-length vector representing the averaged colors of its corresponding images. For GIST, we used the descriptor features first proposed by Oliva and Torralba (2001). This model provides a summary statistic representation of the dominant orientations and spatial frequencies at multiple scales coarsely localized on the image plane. We used spatial bins at 4 cycles per image and 8 orientations at each of 4 spatial scales for a total of 3.072 filter outputs per image. We averaged the GIST descriptors for each of the 16 distinct stimuli in each subordinate category to arrive at a 3072-dimensional representation of each of our 64 subordinates. For wavelet features, we represented each image in our stimulus set as the output of a bank of multi-scale Gabor filters. This type of representation has been used to successfully model the representation in early visual areas (Kay et al., 2008). Each image was converted to gray-scale, downsampled to 128 by 128 pixels, and represented with a bank of Gabor filters at three spatial scales (3, 6, and 11 cycles per image with a luminance-only wavelet that covers the entire image), four orientations (0, 45, 90, and 135°), and two quadrature phases (0 and 90°). An isotropic Gaussian mask was used for each wavelet, with its size relative to spatial frequency, such that each wavelet has a spatial frequency bandwidth of one octave and an orientation bandwidth of 41°. Wavelets were truncated to lie within the borders of the image. Thus, each image is represented by $3*3*2*4+6*6*2*4+11*11*2*4=1328$ total Gabor wavelets. We created the wavelet representation of each of our 64 subordinate categories by averaging over the representation of the 16 distinct images associated with each of them.

2.4.4. Whole-brain searchlight analysis

For each participant's brain, we extracted all gray matter voxels and placed a sphere of radius 4 voxels at every other voxel location (step size: 2 voxels). We excluded all locations where half or more of the voxels in the proposed cube did not overlap with gray matter. For each cube, we computed a local category boundary effect (CBE) for responses to the most typical and the least typical half of our dataset, similar to the analysis procedure described above. We then used these values to identify brain regions where category boundaries were stronger between more typical categories (more typical half CBE > less typical half CBE) and vice versa (more typical half CBE < less typical half CBE). Individual subject results were transformed into group space by aligning to the Talairach atlas and averaging the aligned maps together. To establish statistical significance for our results, we thresholded the group maps for each analysis by using a false discovery rate (FDR) of 0.05, which was determined by computing 1000 simulated group maps, obtained by permuting the category labels without replacement in each voxel cube searchlight.

2.5. Statistical analyses

For all our experiments, we used paired two-tailed t-tests when comparing observed effects against chance and when establishing whether a significant difference exists between two observed effects. We used Kolmogorov–Smirnov tests to establish that no significant deviation from normality exists for the distributions of all effects to which t-tests were applied. All statistical tests were implemented in MATLAB.

3. Results

3.1. Typical exemplars are more neurally similar to category central tendency

Using two separate behavioral experiments (see the Materials and methods section), we established a dataset of eight verified basic level categories (4 natural/animate and 4 man-made/inanimate), each of which comprised eight subordinate level categories normed according to their typicality. Henceforth, we will use the term 'category' to refer to one of our eight basic level categories and the term 'exemplar' to refer to one of our sixty-four subordinate level categories. To investigate whether the family resemblance hypothesis is upheld in visual cortex neural patterns of activation, we scanned participants viewing our sixty-four exemplars (16 visually different images per exemplar, see the Materials and methods section). Since psychological representations of categories are influenced by factors such as task, learning, and attention (Nosofsky, 1992; Love, 2005; Harel et al., 2014), we asked participants to perform a one-back repetition task in the scanner (i.e. no explicit categorization or typicality judgment task) used solely to ensure that they maintained alertness during the experiment. Our analyses focused on object-selective cortex (lateral occipital complex (LOC)) and early visual areas (V1, V2, V3v, hV4).

First, we assessed the intra-class component of the family resemblance hypothesis, namely that more typical exemplars in a category share more features in common with the central category tendency that do atypical exemplars. To test this, within each of our eight categories, we compared how similar (using Pearson's r) the most typical and least typical exemplars were to the central category tendency, defined here by averaging together the neural patterns corresponding to all exemplars in each category. This definition is similar to that of a putative prototype for that category (Sigala and Logothetis, 2002).

Here, we hypothesized that if family resemblance provides a good model for the organization of neural patterns of activation elicited by real-world objects with respect to their typicality, then more typical items should sit closer to the center of this space and hence be more similar to the central category tendency, than the atypical exemplars. Indeed, we found that highly typical exemplars were by far more similar to the category average than less typical exemplars in object-selective cortex (i.e. LOC), but not in early visual areas (Fig. 2; LOC: high > low $t(10) = 3.8$, $p = 0.003$; V1: high > low $t(10) < 1$, $p = 0.491$; V2: high > low $t(10) = 1.3$, $p = 0.228$; V3v: high > low $t(10) < 1$, $p = 0.468$; hV4: high > low $t(10) = 1.2$, $p = 0.261$). Additionally, these results replicate using a version of the analysis where we omitted from the computation of the central tendency the most typical and least typical exemplars (leaving only the six middle-typicality exemplars in each category, see Supplementary Fig. S4). Interpreted differently, an equivalent prediction of the family resemblance hypothesis is that the degree of similarity of each subordinate within a basic category to the most typical subordinate in that category should consistently decrease with the typicality rating given to that particular subordinate. Indeed, we found that this alternative prediction mirrors our results above: similarity is highest between the two most typical subordinates within a basic category and drops successively as typicality for a given subordinate decreases (see Supplementary Fig. S5). Together, these findings show that intra-class structure of real-world categories is consistent with the family resemblance hypothesis in LOC and provides evidence that the representation of object categories shares key properties in common with prototype- and norm-based representations (see e.g. Sigala et al., 2002; Leopold et al., 2006; Abbott et al., 2012).

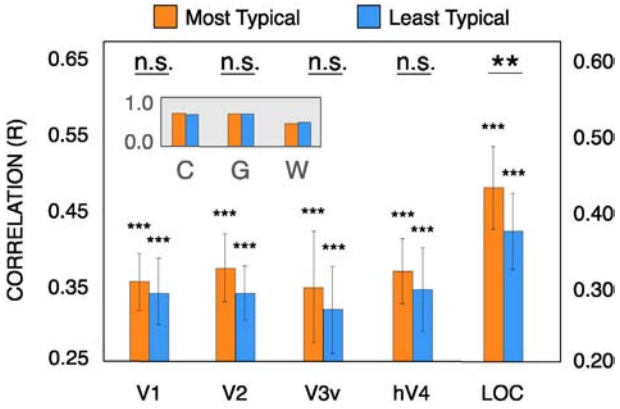

Fig. 2. Typical exemplars are more correlated with category central tendency than less typical exemplars in object-selective cortex. Correlation between category central tendency and most typical exemplar in each category (orange) or least typical exemplar in each category (blue), averaged across all 8 basic level categories. In object-selective cortex (LOC), typical categories are more similar to the average category representation than less typical categories and this effect is not present in early visual areas. (Inset) We performed a similar analysis using the image-level features from our stimulus set: LAB color histograms (C), GIST features (G), and multi-scale Gabor wavelet features (W). All features show similar values for both highly typical and less typical exemplar correlations, with the GIST and wavelet features exhibiting an opposite trend to our LOC results (higher correlation for less typical exemplars). Therefore, low-level stimulus features cannot solely explain our results in object-selective cortex. ***P < 0.001, **P < 0.01, n.s. — not significant. Error bars: 95% confidence interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To show that the effects we observed cannot be explained solely on the basis of the low-level properties of the stimuli themselves, we replicated our similarity analysis using several sets of descriptor features extracted from our images: LAB color histograms, GIST (Oliva and Torralba, 2001), and multi-scale Gabor wavelet features (Kay et al., 2008) (see the Materials and methods section for details on how each of the features was computed). We found that all features show similar numerical correlations between the most typical and least typical exemplars with the central category tendency. Additionally, for GIST and wavelet features, we saw an opposite pattern to our LOC results, namely that correlation with the central category tendency was numerically higher for exemplars ranked as less typical (GIST high $r = 0.86$, low $r = 0.84$; wavelet: high $r = 0.60$, low $r = 0.64$; color: high $r = 0.88$, low $r = 0.84$). For color histograms, a small trend is observed for typical exemplars to be more correlated with the central category tendency, however this trend disappears (and in fact reverses) when excluding the most and the least typical exemplars from the computation of the central category tendency (middle-six exemplars analysis: GIST: high $r = 0.87$, low $r = 0.89$; wavelet: high $r = 0.54$, low $r = 0.60$; color: high $r = 0.91$, low $r = 0.93$; see Supplementary Fig. S4). Overall, this implies that low-level features alone cannot fully account for the pattern of results we observe in object-selective cortex, and further suggests that the human visual system likely constructs (or, at the very least, strongly amplifies) feature descriptions of our visual input that correlate with behavioral typicality judgments later on.

3.2. Typical exemplars exhibit stronger inter-category boundaries

We saw that typicality is correlated with how similar an exemplar is to its central category tendency. Next, we investigated whether typicality affects the second dimension of the family resemblance hypothesis: are typical exemplars more dissimilar to other categories than atypical ones? We hypothesized that if this is the case, then categories defined solely by relatively higher typicality exemplars would be more

distinguishable from one another than categories comprising only less typical exemplars. As such, we split our dataset into to halves, corresponding to the most typical and least typical exemplars from each category. We subsequently computed the category boundary effect (Kriegeskorte et al., 2008; Iordan et al., 2015) for each of the two halves of the dataset as the difference between within-category similarity and between-category similarity, averaged across our eight basic level categories. We predicted that if the family resemblance hypothesis holds, then the category boundary effect would be stronger when computed on the half of the dataset comprising the four most typical exemplars from each category than when computed on the half of the dataset consisting of the least typical four exemplars from each category. Using this measure of how separable categories are in the space of neural patterns of activation, we found that typical exemplars are more easily distinguishable than less typical exemplars in object-selective cortex (Fig. 3; LOC: most typical > least typical, $t(10) = 3.0$, $p = 0.013$). By contrast, typicality does not modulate how separable categories are in the space of neural activations in early visual areas (V1: most typical > least typical, $t(10) < 1$, $p = 0.597$; V2: most typical > least typical, $t(10) = 1.5$, $p = 0.167$; V3v: most typical > least typical, $t(10) = 1.1$, $p = 0.298$; hV4: most typical > least typical, $t(10) = 1.9$, $p = 0.092$).

Analogously to our previous analysis, we asked whether low-level features of our stimulus set are sufficient to explain the pattern of results we observed in object-selective cortex. Accordingly, we computed the category boundary effect on feature descriptors (LAB color histograms, GIST, and multi-scale Gabor wavelet features) extracted from the most typical half and least typical half of our dataset. For all of our feature representations, we found an opposite effect to the one present in LOC: numerically more pronounced category boundaries for the less typical half of our dataset, compared to the most typical half (high vs. low category boundary: color 0.09 vs. 0.14; GIST 0.13 vs. 0.14; wavelet 0.27 vs. 0.32). These results, together with the finding that category boundaries are identical in early visual areas for the two halves of our dataset, provide evidence that it is unlikely that low-level features are directly responsible for the emergence of the typicality effect we observe in object-selective regions. In short, this suggests that typical exemplars become more separated in their neural representation in LOC, and that this effect is not purely driven by the visual appearance of our exemplars and categories, but instead is a direct result of sequential processing along the ventral visual stream.

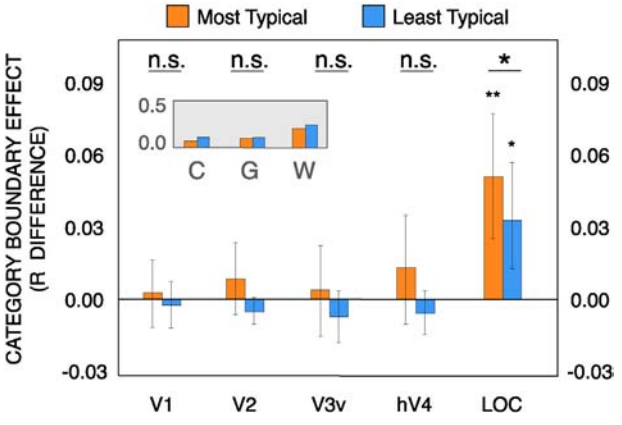

Fig. 3. Category boundaries are stronger for highly typical exemplars in object-selective cortex. Category boundary effect for the two halves of our dataset comprising the most typical 4 exemplars from each category (orange) and the least typical 4 exemplars from each category (blue). In object-selective cortex (LOC), typical exemplars from one category are more distinguishable from exemplars of other categories, an effect not reflected in early visual areas' patterns of activation. (Inset) We performed a similar analysis using the image-level features from our stimulus set: LAB color histograms (C), GIST features (G), and multi-scale Gabor wavelet features (W). All of the feature representations show an opposite trend to that observed in LOC (stronger category boundaries for less typical items) and therefore cannot fully explain our results in object-selective cortex. $**P < 0.01$, $*P < 0.05$, n.s. — not significant. Error bars: 95% confidence interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Finally, the category boundary effect is a compound measure that relies on both within-category similarity (category cohesion) and between-category dissimilarity (category distinctiveness) (Kriegeskorte et al., 2008; Iordan et al., 2015). To investigate the contributions of each of these components of category representation on the strength of the typicality effect we observed, we computed these measures separately for our two halves of the dataset comprising the most and least typical categories, respectively. In all visual areas, we observed no significant differences in cohesion or distinctiveness between the two halves of our dataset (cohesion: LOC: most typical > least typical, $t(10) = 1.7$, $p = 0.120$; V1: most typical > least typical, $t(10) < 1$, $p = 0.564$; V2: most typical > least typical, $t(10) = 1.5$, $p = 0.153$; V3v: most typical > least typical, $t(10) < 1$, $p = 0.631$; hV4: most typical > least typical, $t(10) < 1$, $p = 0.763$; distinctiveness: LOC: most typical > least typical, $t(10) < 1$, $p = 0.736$; V1: most typical > least typical, $t(10) < 1$, $p = 0.735$; V2: most typical > least typical, $t(10) < 1$, $p = 0.537$; V3v: most typical > least typical, $t(10) < 1$, $p = 0.760$; hV4: most typical > least typical, $t(10) = -1.2$, $p = 0.247$). Considering our main finding that a significant difference exists between category boundaries elicited by more and less typical exemplars in LOC, the lack of a significant effect for cohesion and distinctiveness suggests that neither within-category similarity, nor between-category similarity differences drive our effects on their own, but rather it is their combined effect (difference) that separates typical and atypical exemplars in this brain region.

An analogous prediction of this second aspect of the family resemblance hypothesis indicates that if typical subordinates are indeed more separable from other categories, then they should sit farther from a putative fixed category boundary between two basic categories compared to less typical categories. Indeed, a separate analysis that defined fixed support-vector-machine (SVM) boundaries between every pair of basic categories indicated that, on average, the most typical four subordinates in each category exhibited larger distances to their corresponding boundary than the four least typical subordinates in LOC, but not in early visual regions (Supplementary Fig. S6).

Overall, our findings provide strong evidence in favor of the neural plausibility of the family resemblance hypothesis in LOC. In this brain region, typical exemplars are more similar to the average category representation and are more separable (as conferred by their larger category boundary effect) across categories than atypical exemplars, which suggests that typicality exerts a measurable and consistent modulatory effect on the nature of the distributed patterns of neural representation of real-world object categories in object-selective cortex.

3.3. Whole-brain analysis

So far, we have limited our analyses to functionally defined cortical areas. However, it may be the case that activity in other brain areas beyond our pre-selected ROIs may favor the representation or dissociation of typical and atypical exemplars from the same category. To investigate this hypothesis, we performed a whole-brain searchlight analysis (Kriegeskorte et al., 2006) where we computed the category boundary effect for the most typical half of the dataset and the least typical half of the dataset for equally spaced spheres of voxels tiling the entire gray matter surface of our participants' brains. This analysis identifies brain regions where typicality organizes the neural representation space according to the family resemblance hypothesis (typical exemplars more similar to central category tendency, while maximizing distance to other categories). More interestingly, by performing the reverse contrast, we may also uncover brain regions where the opposite is true: since we know that even atypical exemplars are still identified as members of their respective categories, it is likely that computations exist which are meant to ensure differentiation between these

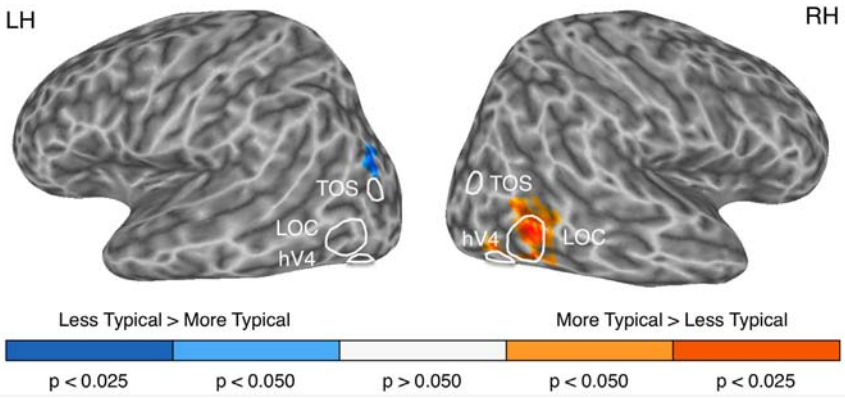

Fig. 4. Whole-brain searchlight analysis uncovers brain regions where category boundaries are stronger between most typical and least typical exemplars. We performed a whole-brain searchlight analysis where we computed the difference between the category boundary effects obtained for the most typical half of our dataset and the least typical half of our dataset. Figure shows group map results, corrected for multiple comparisons using an FDR measure (see the Materials and methods section for details). Regions shown in orange (right LOC, right hV4) showed a significant effect of typicality: highly typical exemplars were more distinguishable from exemplars of other categories. Conversely, regions shown in blue (left cIPL) showed the opposite trend: less typical exemplars were more easily distinguishable form members of other categories. This cortical region has been previously implicated in category learning (Zeithamova et al., 2008) and contextual processing (Konen and Kastner, 2008), which suggests the possibility that it may aid in the categorization of atypical items, perhaps through mediating contextual facilitation of recognition. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

exemplars and thus enable correct assignment into their purported categories.

Consistent with our previous ROI results, we found that typicality modulates the strength of category distinctions in the right LOC and to a lesser extent in a region adjacent to the right hV4 (Fig. 4, right). This finding indicates that, indeed, typicality modulates representation of object categories in object-selective cortex and that this effect is strongest in this region, not simply a late vs. early visual cortex difference in representation.

Interestingly, we also uncovered an advantage for neural patterns of activation distinguishing best between atypical exemplars, compared to highly typical exemplars, in the caudal inferior parietal lobule (cIPL; Fig. 4, left). This region has been previously implicated in contextual processing (Konen and Kastner, 2008) and category learning (Zeithamova et al., 2008), which raises the possibility that enhanced category boundaries for atypical categories here may be due to additional or specialized processing required to disambiguate between less typical exemplars and subsequently assign them a correct category label.

Taken together, our results suggest that typicality is linked to the neural representation of object categories across several brain regions, with its effects extending to both intra-class and inter-class organization. Our results provide neural confirmation for both predictions of the family resemblance hypothesis in object-selective cortex (Rosch and Mervis, 1975) and, furthermore, we provide the first evidence that typicality provides a concrete dimension of neural organization for real-world object categories in both object-selective cortex (LOC) and cIPL, but outside of early visual cortex, which further suggests that this representation is not directly reflected in image features describing natural input, but rather built by the visual system at an intermediate processing stage.

4. Discussion

Typicality is a ubiquitous, yet often overlooked property of virtually all objects we interact with in our visual environment. Despite well-studied and long-standing behavioral effects associated with typicality, such as increased speed of recognition and lower error rates for identifying the category membership of more typical items (Posner and Keele, 1968; Rosch, 1973; Rosch and Mervis, 1975), little is known about how typicality relates to the neural representation of objects from the same category. Our work is the first to address this fundamental question using a large array of real-world stimuli. As such, we provide the first neural test of the predictions of the family resemblance hypothesis for real-world object categories: namely, that highly typical exemplars share most features in common with other members of their category (e.g. 'Golden retriever' is a highly representative dog), while simultaneously sharing the fewest features in common with other exemplars from semantically-related categories (e.g. Golden retrievers share fewer features with cats than less typical exemplars such as Chihuahuas). Using several similarity-based multivariate pattern analyses, which make no explicit assumptions regarding the nature of the neural feature space in which objects are represented, we found that this conception of category structure describes the organization of neural patterns better in object-selective regions than in early visual areas of the brain. Coupled with the fact that this representation is not directly reflected in image features describing natural input, these data suggest that such a representation is not given in the input, but rather built by the visual system at an intermediate processing stage. In the current set of experiments, we exclusively investigated how typicality affects the neural representation of a set of carefully normed, hierarchically organized object categories. While there is no reason to believe that a separate collection of categories (i.e. one not possessing a taxonomic relationship) would behave differently within the context of neural typicality measures as exemplified in our results, such an experiment remains an interesting question for future work.

The neural basis of typicality has been previously investigated almost exclusively using learning paradigms over artificially constructed category spaces (see e.g. Aizenstein et al., 2000; Sigala et al., 2002; Sigala and Logothetis, 2002; Davis and Poldrack, 2014). One of the main advantages of using artificial categories is the tremendous degree of control one possesses over the instantiation of the feature space, as well as the stimuli themselves. Additionally, synthetic category spaces remove all potential confounds related to object properties that may be directly linked to typicality itself, such as familiarity, discriminability, and expertise. Nevertheless, these idealized and impoverished spaces not only noticeably lack the complexity of visual stimuli we encounter in our everyday environment, but participants' experience with them is necessarily more limited, leaving open the question as to what degree such findings generalize to the real world and to categories that are overlearned. By testing the predictions of the family resemblance hypothesis on real-world categories directly, our current experiment provides long overdue concrete evidence for a typicality-based organization of the neural representation space for such categories in human visual cortex. In our experiment, we not only found that highly typical objects generate stronger category boundaries in object-selective cortex, but we also uncovered the first evidence for a brain region where the opposite is true: in the caudate inferior parietal lobule (cIPL), we see atypical exemplars becoming more differentiated by neural patterns

of activity than their highly typical counterparts. This region is superior to the trans-occipital sulcus and the functionally defined scene-selective region TOS (or OPA) (Grill-Spector, 2003; Dilks et al., 2013), likely overlapping with functionally defined area IPS0 (Silver and Kastner, 2009). A representation of objects is known to exist in posterior parietal cortex (PPC), independent of action planning, and this cortical region has been shown to exhibit adaptation to object properties, including shape and size (Konen and Kastner, 2008). Furthermore, the PPC has also been implicated in the learning of new categories (Zeithamova et al., 2008), in the recall of words and objects, provided that the stimuli are associated with strong memory of source context (Johnson and Rugg, 2007; Peters et al., 2009; Vilberg and Rugg, 2009, 2012), as well as in the representation of perceptual decision variables (Heekeren et al., 2006; Tosoni et al., 2008). Taken together, these findings raise the possibility that this cortical region may aid in the categorization of atypical items, perhaps through mediating contextual facilitation of recognition. Intuitively, processing category boundaries both in terms of typical and atypical exemplars is potentially necessary for arriving at a unified percept of a category: to recognize a 'dog' in our visual interaction with the world, our brain must understand both what a dog usually looks like (typicality), as well as what degree of deviation from this representation should place our percept outside of that particular category.

Nevertheless, caution is necessary in interpreting these results, especially in dorsal stream regions: given that typicality is a subjective measure that subsumes multiple dimensions and features of object categories (including e.g. frequency of occurrence in the world and familiarity with such objects), the possibility exists that our findings may have been influenced by differences in the allocation of attentional resources across such dimensions (e.g. if participants paid more attention to blocks containing less familiar subordinate categories). However, our searchlight analysis identified regions where the category boundary effect (computed via the similarity of multi-voxel patterns) differs consistently between typical and atypical members of our categories, which indicates the presence of discriminable category information in these brain regions. Thus, if attention plays a role in our findings, then it would necessarily have to be operating on the category representations themselves, bringing within category members closer in neural space and pulling between category members apart. Additionally, previous work has shown that two parallel and hierarchically organized neural systems for object representation exist along the ventral and dorsal pathways (Konen and Kastner, 2008; Wurm and Lingnau, 2015, Vaziri-Pashkam and Xu, 2015; Braunlich and Seger, 2016) and our results in cIPL are consistent with such an account.

Recent work has shown that distance from an inferred category boundary constructed from patterns of neural activation in human inferotemporal cortex can be used to successfully predict behavioral categorization (Carlson et al., 2013; Ritchie et al., 2015). This distance-based model of category representation is consistent with our results in LOC, where we show that category boundaries are stronger between highly typical exemplars than between less typical exemplars, with the latter sitting farther from the category central tendency. Relatedly, many distance metrics have been previously employed for characterizing the similarity of neural patterns of activity in human visual cortex in general, and typicality in particular, ranging from overall cortical activity level (Leopold et al., 2006; Park et al. 2015) to Pearson correlation (e.g. Haxby et al., 2001; Davis and Poldrack, 2014; Iordan et al., 2015) and Euclidean or city block distance (Ashby and Maddox, 1993; Sigala et al., 2002). Of these, we chose to focus on Pearson correlation as a straightforward, scale-invariant measure of similarity of neural patterns (Davis et al., 2014). This is especially relevant, given that we perform a large-scale experiment using 64 real-world categories and prior evidence has shown that objects from different categories have the potential to elicit consistently different univariate activity profiles both within and between brain regions (e.g. animate vs. inanimate categories (Connolly et al., 2012; Konkle and Caramazza, 2013), small vs. big objects (Konkle and Oliva, 2012; Konkle and Caramazza, 2013)). Moreover, our decision is consistent with analyses used in many recent experiments investigating the underpinnings of object categorization and typicality in humans and non-human primates (e.g. Haxby et al., 2001; Kriegeskorte et al., 2008; Connolly et al., 2012; Davis and Poldrack, 2014; Iordan et al., 2015).

Several cognitive theories have been proposed that suggest that we may expect real-world object categories to have a strong prototype-dominated cortical representation (Nosofsky, 1991; Ashby and Maddox, 1993), with typical exemplars closer in neural distance to the basic level prototype (category central tendency) and less typical exemplars generating a more distinct neural pattern of activation (i.e. larger neural distance from prototype). Indeed, previous work involving artificially constructed face stimuli suggests that both feature-based and neural distance from a category central tendency are usually correlated with perceived typicality (Leopold et al., 2001; Sigala et al., 2002; Leopold et al., 2006; Davis and Poldrack, 2014). Prototype theory is typically contrasted with exemplar theory, which proposes that we represent categories with respect to several emblematic exemplars (or perhaps all exemplars) in each category, which serve to map that particular category's representational space (Nosofsky, 1986, 1991; Ashby and Maddox, 1993). This theory has also received some support; recent work has shown that exemplar models explain a comparable amount of variance in human performance on category generalization and prediction tasks (Abbott et al., 2012) and even surpass prototype models in performance using data from humans and monkeys categorizing cartoon depictions of faces and fish (Sigala et al., 2002). In our work, we find brain areas that separately emphasize characteristics from both of these putative representational models, raising the possibility that the human brain may use both strategies for forming categories. First, we show that, in object-selective regions, typical categories are closer to the central category tendency and category boundaries are sharpened between typical and atypical exemplars, a finding that is consistent with the family resemblance hypothesis, as well as with a prototype-based encoding of category structure (but see (Ashby and Maddox, 1993) for an alternate explanation of how exemplar theory may also account for such a prediction). Conversely, we also find that atypical exemplars exhibit stronger category boundaries in cIPL. One potential explanation for this finding is that real-world categories, especially due to their inherent intra-class complexity, may not be fully or accurately captured by a single prototype per category. Thus, while a prototype representation would imply that the intra-class distribution of subordinate categories within a basic is less important compared to the location of the category central tendency (i.e. prototype), by contrast an exemplar representation would predict a much heavier reliance on less typical subordinates for differentiating between basic categories, which may be the case in cIPL. Taken together, these two contrasting patterns of results suggest that the human brain may, in fact, use both exemplar and prototype models to structure category representations, albeit in different brain regions. Such a position could reconcile the seemingly contradictory behavioral and modeling results that have yet to eliminate either model as the sole framework for intra-category organization (see e.g. Sigala et al., 2002 and Leopold et al., 2006). Critically, our results provide clear evidence that LOC and cIPL are strong candidates for future investigations attempting to elucidate the contributions of these individual models in explaining the eventual emergence of perceptual typicality.

Over the past two decades, evidence has been uncovered for specific cortical regions selective for broad stimulus classes such as faces, scenes, objects, and bodies (Malach et al., 1995; Kanwisher et al., 1997; Epstein and Kanwisher, 1998; Downing et al., 2001), as well as organizational principles corresponding to broad attribute dimensions, including animacy (Chao et al., 1999; Kriegeskorte et al., 2008; Connolly et al., 2012; Konkle and Caramazza, 2013) and real-world object size (Konkle and Oliva, 2012; Konkle and Caramazza, 2013). Furthermore, many studies have demonstrated that category information is recoverable from distributed representations (Haxby et al., 2001; Cox and Savoy, 2003;

Haynes and Rees, 2005; Eger et al., 2008; Huth et al., 2012), yet what constitutes a category representation in the high-dimensional space of neural patterns of activity is still poorly understood. Here, we show that perceived typicality, a high-level cognitive property of objects, directly modulates the representation of exemplars and categories fairly early in visual processing. Our results raise the possibility that the same theoretical principles that guide the cognitive formation of categories (cognitive usefulness and feature correlation constraints present in the environment (Rosch et al., 1976)) may, in fact, fundamentally and sequentially guide the processing of visual input from its very early cortical stages. Indeed, previous work from our lab has already shown that this early link to cognition also holds for hierarchical organization of category structure, whose influence on the organization of neural patterns becomes apparent as early as lateral occipito-temporal cortex (Iordan et al., 2015). In the process of building category representations, the inclusion of such principles would improve the utility and flexibility of eventually generated categories by emphasizing better boundaries between them and by allowing distinctions between individual exemplars and multiple levels of generality to emerge gradually from the neural representation. Furthermore, such principles constitute important signposts for recent work whose goal is to map the layers of deep learning models for visual categorization onto successive stages of the ventral visual hierarchy (Cadieu et al., 2014; Yamins et al., 2014; Yamins and DiCarlo, 2016). Most such computational models include few, if any, high-level cognitive constrains on their internal representation aside from categorization itself as an end-goal. Moving forward, we argue that attempts to build models of visual processing that more accurately mirror the human visual processing hierarchy would benefit from incorporating (either explicitly or at a verification stage) other high-level properties such as typicality, which we have presently identified as having a measurable impact on the feature spaces of visual regions strongly involved in object and category recognition (e.g. LOC).

Together, these findings solidify our understanding of how we define and describe boundaries between category representations in the brain, and moreover, put forward a new hypothesis for the organization and goals of intermediate visual processing: it is not simply focused on isolating and identifying primitives such as shapes, objects, or scenes, and their interplay, but also on employing cognitively relevant principles of category organization (of which typicality and hierarchical organization are two examples) to directly guide the development of the neural representation, for the ensuing purpose of improved and more flexible categorization, action, and cognition.

Competing interests

The authors declare no competing financial interest.

Author contributions

M.C.I., M.R.G., D.M.B., and L.F.-F. designed the experiments. M.R.G. collected the behavioral data. M.C.I. and M.R.G. analyzed the behavioral data and collected the fMRI data. M.C.I. analyzed the fMRI data. M.C.I., M.R.G., D.M.B., and L.F.-F. wrote the manuscript.

Funding acknowledgments

This work was funded by the William R. Hewlett Stanford Graduate Fellowship (to M.C.I.), the William and Adeline Hendess Phi Beta Kappa Graduate Fellowship (to M.C.I.), an NRSA Grant from the National Eye Institute NEIF32EY019815 (to M.R.G.), and a National Institutes of Health Grant 1 R01 EY019429 (to D.M.B and L.F.-F.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.04.012.

References

Abbott, J.T., Austerweil, J.L., Griffiths, T.L., 2012. Constructing a hypothesis space from the web for large-scale Bayesian word learning. In: Miyake, N., Peebles, D., Cooper, R.P. (Eds.), Proceedings of the 34th Annual Conference of the Cognitive Science Society. Cognitive Science Society, Austin (TX), pp. 54–59.
Aizenstein, H.J., MacDonald, A.W., Stenger, V.A., Nebes, R.D., Larson, J.K., Ursu, S., Carter, C.S., 2000. Complementary category learning systems identified using event-related functional MRI. J. Cogn. Neurosci. 12 (6), 977–987.
Ashby, F.G., Maddox, W.T., 1993. Relations between prototype, exemplar, and decision bound models of categorization. J. Math. Psychol. 37, 372–400.
Ashby, F.G., Maddox, W.T., 2005. Human category learning. Annu. Rev. Psychol. 56, 149–178.
Biederman, I., 1987. Recognition by components: a theory of human image understanding. Psychol. Rev. 94 (2), 115–117.
Brainard, D.H., 1997. The Psychophysics Toolbox. Spat. Vis. 10 (4), 433–436.
Braunlich, K., Seger, C.A., 2016. Categorical evidence, confidence, and urgency during probabilistic categorization. NeuroImage 125, 941–952.
Cadieu, C.F., Hong, H., Yamins, D.L.K., Pinto, N., Ardila, D., Solomon, E.A., Majaj, N.J., DiCarlo, J.J., 2014. Deep neural networks rival the representation of primate IT cortex for core visual object recognition. PLoS Comput. Biol. 10 (12), e1003963.
Carlson, T.A., Simmons, R.A., Kriegeskorte, N., Slevc, L.R., 2013. The emergence of semantic meaning in the ventral temporal pathway. J. Cogn. Neurosci. 26 (1), 120–131.
Chao, L.L., Haxby, J.V., Martin, A., 1999. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat. Neurosci. 2 (10), 913–919.
Clarke, A., Tyler, L.K., 2014. Object-specific semantic coding in human perirhinal cortex. J. Neurosci. 34 (14), 4766–4775.
Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.C., Abdi, H., Haxby, J.V., 2012. The representation of biological classes in the human brain. J. Neurosci. 32 (8), 2608–2618.
Cox, D.D., Savoy, R.L., 2003. Functional magnetic resonance imaging (fMRI) 'brain reading': detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage 19 (2), 261–270.
Davis, T., Poldrack, R.A., 2014. Quantifying the internal structure of categories using a neural typicality measure. Cereb. Cortex 26, 1–18.
Davis, T., Love, B.C., Preston, A.R., 2012a. Learning the exception to the rule: model-based fMRI reveals specialized representations for surprising category members. Cereb. Cortex 22, 260–273.
Davis, T., Love, B.C., Preston, A.R., 2012b. Striatal and hippocampal entropy and recognition signals in category learning: simultaneous processes revealed by model-based fMRI. J. Exp. Psychol. Learn. Mem. Cogn. 38 (4), 821–839.
Davis, T., LaRocque, K.F., Mumford, J.A., Norman, K.A., Wagner, A.D., Poldrack, R.A., 2014. What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. NeuroImage 97, 271–283.
Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. ImageNet: a large-scale hierarchical image database. Conf. Comput. Vis. Pattern Recognit. 248–255.
Dilks, D.D., Julian, J.B., Paunov, A.M., Kanwisher, N., 2013. The occipital place area is causally and selectively involved in scene perception. J. Neurosci. 33 (4), 1331–1336.
Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. Science 293, 2470–2473.
Eger, E., Ashburner, J., Haynes, J.D., Dolan, R.J., Rees, G., 2008. fMRI activity patterns in human LOC carry information about object exemplars within category. J. Cogn. Neurosci. 20 (2), 356–370.
Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. Nature 392, 598–601.
Golarai, G., Ghahremani, D.G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J.L., Gabrieli, J.D.E., Grill-Spector, K., 2007. Differential development of high-level visual cortex correlates with category-specific recognition memory. Nat. Neurosci. 10 (4), 512–522.
Grill-Spector, K., 2003. The neural basis of object perception. Curr. Opin. Neurobiol. 13, 159–166.
Harel, A., Kravitz, D.J., Baker, C.I., 2014. Task context impacts visual object processing differentially across the cortex. Proc. Natl. Acad. Sci. U. S. A. 111 (10), E962–E971.
Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430.
Haynes, J.D., Rees, G., 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat. Neurosci. 8 (5), 686–691.
Heekeren, H.R., Marrett, S., Ruff, D.A., Bandettini, P.A., Ungerleider, L.G., 2006. Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality. Proc. Natl. Acad. Sci. U. S. A. 103 (26), 10023–10028.
Huth, A.G., Nishimoto, S., Vu, A.T., Gallant, J.L., 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76 (6), 1210–1224.
Iordan, M.C., Greene, M.R., Beck, D.M., Fei-Fei, L., 2015. Basic level category structure emerges gradually across human ventral visual cortex. J. Cogn. Neurosci. 27 (7), 1–29.
Johnson, J.D., Rugg, M.D., 2007. Recollection and the reinstatement of encoding-related cortical activity. Cereb. Cortex 17, 2507–2515.
Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17 (11), 4302–4311.
Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L., 2008. Identifying natural images from human brain activity. Nature 452, 352–355.
Konen, C.S., Kastner, S., 2008. Two hierarchically organized neural systems for object information in human visual cortex. Nat. Neurosci. 11 (2), 224–231.

Konkle, T., Caramazza, A., 2013. Tripartite organization of the ventral stream by animacy and object size. J. Neurosci. 33 (25), 10235–10242.
Konkle, T., Oliva, A., 2012. A real-world size organization of object responses in occipitotemporal cortex. Neuron 74 (6), 1114–1124.
Kriegeskorte, N., Goebel, R., Bandettini, P.A., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U. S. A. 103 (10), 3863–3868.
Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., Bandettini, P.A., 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60 (6), 1126–1141.
Leopold, D.A., O'Toole, A.J., Vetter, T., Blanz, V., 2001. Prototype-referenced shape encoding revealed by high-level aftereffects. Nat. Neurosci. 4 (1), 89–94.
Leopold, D.A., Bondar, I.V., Giese, M.A., 2006. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. Nature 442, 572–575.
Love, B.C., 2005. Environment and goals jointly direct category acquisition. Curr. Dir. Psychol. Sci. 14, 195–199.
Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc. Natl. Acad. Sci. U. S. A. 92 (18), 8135–8139.
Minda, J.P., Smith, J.D., 2002. Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. J. Exp. Psychol. Learn. Mem. Cogn. 28 (2), 275–292.
Nosofsky, R.M., 1986. Attention, similarity, and the identification-categorization relationship. J. Exp. Psychol. Gen. 115 (1), 39–61.
Nosofsky, R.M., 1991. Typicality in logically defined categories: exemplar-similarity versus rule instantiation. Mem. Cogn. 19 (2), 131–150.
Nosofsky, R.M., 1992. Similarity scaling and cognitive process models. Annu. Rev. Psychol. 43, 25–53.
Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. J. Comput. Vis. 42 (3), 145–175.
Park, S., Konkle, T., Oliva, A., 2015. Parametric coding of the size and clutter of natural scenes in the human brain. Cerebral Cortex 25 (7), 1792–1805.
Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. 10 (4), 437–442.
Peters, J., Daum, I., Gizewski, E., Forsting, M., Suchan, B., 2009. Associations evoked during memory encoding recruit the context-network. Hippocampus 19, 141–151.
Posner, M., Keele, S., 1968. On the genesis of abstract ideas. J. Exp. Psychol. 77 (3), 353–363.
Ritchie, J.B., Tovar, D.A., Carlson, T.A., 2015. Emerging object representations in the visual system predict reaction times for categorization. PLoS Comput. Biol. 11 (6), e1004316.
Rosch, E.H., 1973. On the internal structure of perceptual and semantic categories. In: Moore, T.E. (Ed.), Cognitive Development and the Acquisition of Language. Academic Press, Oxford (UK) (p. xii, 308).
Rosch, E., Mervis, C.B., 1975. Family resemblances: studies in the internal structure of categories. Cogn. Psychol. 7, 573–605.
Rosch, E., Mervis, B., Gray, W.D., Johnson, D.M., Boyes-Braem, P., 1976. Basic objects in natural categories. Cogn. Psychol. 8, 382–439.
Sayres, R., Grill-Spector, K., 2008. Relating retinotopic and object-selective responses in human lateral occipital cortex. J. Neurophysiol. 100 (1), 249–267.
Sigala, N., Logothetis, N.K., 2002. Visual categorization shapes feature selectivity in the primate temporal cortex. Nature 415, 318–320.
Sigala, N., Gabbiani, F., Logothetis, N.K., 2002. Visual categorization and object representation in monkeys and humans. J. Cogn. Neurosci. 14, 187–198.
Silver, M.A., Kastner, S., 2009. Topographic maps in human frontal and parietal cortex. Trends Cogn. Sci. 13, 488–495.
Tosoni, A., Galati, G., Romani, G.L., Corbetta, M., 2008. Sensory-motor mechanisms in human parietal cortex underlie arbitrary visual decisions. Nat. Neurosci. 11 (12), 1446–1453.
Vaziri-Pashkam, M., Xu, Y., 2015. Object representations in human parietal and occipitotemporal cortices: similarities and differences. J. Vis. 15 (12), 374.
Vilberg, K.L., Rugg, M.D., 2009. Functional significance of retrieval-related activity in lateral parietal cortex: evidence from fMRI and ERPs. Hum. Brain Mapp. 30, 1490–1501.
Vilberg, K.L., Rugg, M.D., 2012. The neural correlates of recollection: transient versus sustained fMRI effects. J. Neurosci. 32 (45), 15679–15687.
Walther, D.B., Caddigan, E., Fei-Fei, L., Beck, D.M., 2009. Natural scene categories revealed in distributed patterns of activity in the human brain. J. Neurosci. 29 (34), 10573–10581.
Wurm, M.F., Lingnau, A., 2015. Decoding actions at different levels of abstraction. J. Neurosci. 35 (20), 7727–7735.
Yamins, D.L.K., DiCarlo, J.J., 2016. Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. 19 (3), 356–365.
Yamins, D.L.K., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., DiCarlo, J.J., 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. U. S. A. 111 (23), 8619–8624.
Zeithamova, D., Maddox, W.T., Schnyer, D.M., 2008. Dissociable prototype learning systems: evidence from brain imaging and behavior. J. Neurosci. 28 (49), 13194–13201.