

Creating Concepts from Converging Features in Human Cortex

Marc N. Coutanche and Sharon L. Thompson-Schill

Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA

Address correspondence to Dr Marc Coutanche, Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104, USA. Email: coumarc@psych.upenn.edu

To make sense of the world around us, our brain must remember the overlapping features of millions of objects. Crucially, it must also represent each object's unique feature-convergence. Some theories propose that an integration area (or "convergence zone") binds together separate features. We report an investigation of our knowledge of objects' features and identity, and the link between them. We used functional magnetic resonance imaging to record neural activity, as humans attempted to detect a cued fruit or vegetable in visual noise. Crucially, we analyzed brain activity before a fruit or vegetable was present, allowing us to interrogate top-down activity. We found that pattern-classification algorithms could be used to decode the detection target's identity in the left anterior temporal lobe (ATL), its shape in lateral occipital cortex, and its color in right V4. A novel decoding-dependency analysis revealed that identity information in left ATL was specifically predicted by the temporal convergence of shape and color codes in early visual regions. People with stronger feature-and-identity dependencies had more similar top-down and bottom-up activity patterns. These results fulfill three key requirements for a neural convergence zone: a convergence result (object identity), ingredients (color and shape), and the link between them.

Keywords: anterior temporal lobe, convergence zone, integration, objects, semantic memory

Introduction

We encounter millions of objects during our lifetime that we recognize effortlessly. We know that a lime is green, round, and tart, whereas a carrot is orange, elongated, and sweet, helping us to never confuse the wedge on our margarita glass with our rabbit's favorite treat. One property (feature) alone is typically insufficient: Celery can also be green; tangerines are orange. Instead, we use the unique convergence of features that defines an object. How does our brain bind these sensorimotor features to form a unique memory representation?

One set of theories proposes that knowledge of objects resides in the very sensorimotor cortices that process their features during perception or use (Martin 2007; Kiefer and Pulvermüller 2012). Under this framework, our knowledge of a concept arises from direct connections between sensorimotor regions. Another set of theories suggest that objects become represented in one or more integration areas (Simmons and Barsalou 2003; Patterson et al. 2007). These "integration" theories differ in their details. An early proposal suggested that various convergence zones hold binding codes that link distinct "feature fragments" in sensory cortex (Damasio 1989; Simmons and Barsalou 2003; Meyer and Damasio 2009). More recent integration proposals suggest that a single "hub" brings together known sensory information into a representation of a concept (Lambon Ralph 2014). Specifically, the "hub-and-spoke"

model proposes that while sensory and verbal information is processed in modality-specific regions, a hub, based in the anterior temporal lobe (ATL), contains a high-dimensional modality-independent semantic space that allows computations to be based on semantic information rather than purely sensory similarities (Lambon Ralph 2014). This is analogous to a "hidden layer" in neural network models, which enables computation of nonlinear relationships between the information coded in sensory layers (Rogers et al. 2004; Lambon Ralph et al. 2010).

A broad range of methods point to the ATL as a possible location for a hub. Semantic dementia patients with ATL atrophy demonstrate conceptual impairments (Hodges et al. 1992; Rogers et al. 2007). These deficits cross modalities (Bozeat et al. 2000; Luzzi et al. 2007), reflecting the proposed transmodal nature of ATL representations (Patterson et al. 2007; Lambon Ralph et al. 2010). In further support, temporarily interfering with ATL activity in healthy participants using transcranial magnetic stimulation (TMS) produces semantic impairments across a range of tasks with words and pictures (Pobric et al. 2007, 2010a, 2010b; Ishibashi et al. 2011). It has recently been suggested that information becomes increasingly higher order and transmodal in more rostral ATL areas (Binney et al. 2012). Neuroimaging investigations have differed in whether significant ATL activation is detected during semantic processing, in large-part due to signal loss in this region (Visser, Jefferies et al. 2010). The region is more likely to be identified after correcting for signal distortions (Visser, Embleton et al. 2010; Visser and Lambon Ralph 2011; Visser, Jefferies et al. 2012). Peelen and Caramazza (2012) recently reported that the ATL encodes two conceptual dimensions of viewed objects: how an object is used, and where it is typically kept (kitchen or garage). In contrast, the similarity of activity patterns in posterior occipitotemporal cortex was more closely related to perceptual similarity (although this was also likely influenced by the concurrent visual presentation of objects).

While distinct integration theories differ in their details, they share the hypothesis that featural information is integrated in at least one cortical location (which, following Damasio's terminology, we refer to as a convergence zone for the remainder of the manuscript). A putative convergence zone should show certain characteristics that are testable using novel fMRI analysis techniques. Specifically, the convergence account leads to three key predictions: 1) Thinking about an object should evoke brain activity representing its specific identity in a purported convergence zone (the result of convergence). 2) Retrieving an object from memory should be accompanied by neural information about its characteristic features within specialized regions (the substrates for convergence). The specificity of these feature fragments could range from general shape processing to specifying a sphere rather than a cube (the strictest form of feature fragment). 3) Convergence success

should be linked to the simultaneous presence of convergence substrates: specifically, activation of “convergence zones” would produce synchronous activity in separate cortical sites presumed to contain feature fragments related to the convergence zone” (Damasio 1989, p. 56). We developed a novel analysis to test relationships between different types of pattern information, allowing us to detect a relationship between feature fragments and their convergence into identity in the human brain.

In the present study, we employ a task that engages top-down influences without visual information on-screen, allowing us to investigate retrieved object knowledge. We examine memory-driven activity patterns for fruits and vegetables that vary orthogonally by color, shape, and identity. Our data support the three predictions of a convergence zone, including a link between specific features in visual cortex and object identity in the left ATL.

Materials and Methods

Subjects

Data from 11 participants (3 females, 18–35 years old) are analyzed (a 12th participant’s fMRI data were not analyzed due to abnormal behavioral responses during the task). All participants were right-handed with normal or corrected-to-normal vision and reported no history of neurological problems. Participants provided written informed consent and received monetary compensation for their participation. The human subjects review board at the University of Pennsylvania approved all experimental procedures.

Magnetic Resonance Imaging Acquisition

Subjects were scanned with a 3-T Siemens Trio system equipped with an eight-channel head coil and foam padding for stabilizing the head. T_1 -weighted anatomical images were acquired at the beginning of each session (repetition time [TR] = 1620 ms, TE = 3 ms, TI = 950 ms, voxel size = $0.977 \times 0.977 \times 1.000$ mm). T_2^* -weighted scans recorded blood oxygenation level-dependent (BOLD) response using interleaved gradient-echo EPI (TR = 3000 ms, TE = 30 ms, field of view = 19.2×19.2 cm, voxel size = $3.0 \times 3.0 \times 3.0$ mm, 42 slices).

Experimental Procedure

Prior to fMRI scanning, participants completed a behavioral staircasing procedure to determine the level of visual noise that was later applied to images presented during the fMRI scan. This ensured that the in-scan detection task would be challenging enough to engage each subject. On each trial of this staircasing behavioral task, a MATLAB script presented subjects with an image of a fruit (bananas and tomatoes, 2 fruits not used for the primary task) with an overlaid field of Gaussian visual noise. Subjects indicated with a button-press if they could identify the fruit. After each behavioral response, the script increased or decreased the variance of the noise that was added to the next image, to bring the subject’s final detection level to 75% by the end of the procedure. The variance giving this detection level was then applied in the same way to the images used in each participant’s scan. The participants’ in-scan detection accuracies were very close to this level, suggesting the staircasing was successful (M hit rate = 74%, SD = 14%; M false alarm rate (FAR) to pure noise = 7%; FAR to foils-in-noise = 35%).

At the beginning of the scanning session, participants passively viewed images of exemplars of the 4 types of fruit and vegetables (carrots, celery, limes, and tangerines) that would later act as targets, centrally placed on a white background. During this run, blocks of 6 images of each type of fruit and vegetable were presented in a random order, with each image shown for 3 s. Blocks were separated by 12 s of fixation. In total, 12 exemplars of each type of fruit and vegetable (later hidden in the detection task) were presented, split across 2 blocks.

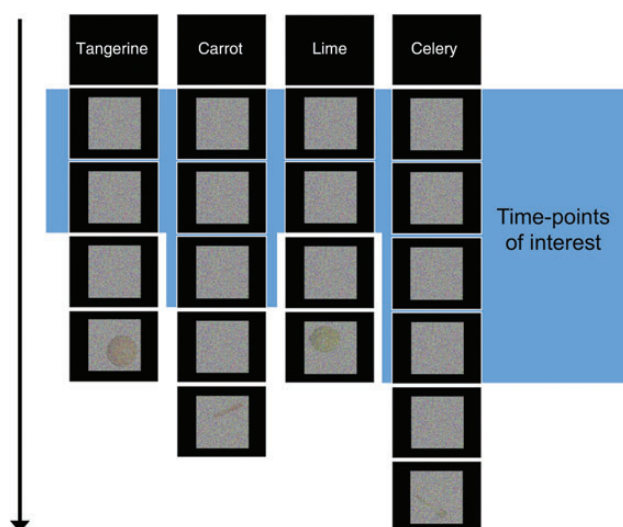


Figure 1. Experimental design. Participants were presented with cues of items to detect, followed by blocks of visual noise. Each block ended with an actual image embedded in noise, at a threshold that was determined for each participant before their scan (shown here at a low threshold for visualization purposes). Blocks contained an unpredictable amount of pure noise and occasionally ended with an incorrect (noncued) fruit or vegetable to keep participants on task. The objects in the final trial are displayed here in each corner although they could appear in any corner in the actual experiment. Every block ended with a unique instance of that kind of fruit or vegetable (e.g., no particular tangerine appeared more than once). Data associated with the last noise time-point (after accounting for the hemodynamic lag) were discarded to ensure that the signal-onset from viewing the image-in-noise did not influence the analyzed data.

During the next 4 scanning runs, participants were instructed to respond with a button-press when they detected a cued fruit or vegetable within visual noise. Word cues were presented to indicate a type of fruit or vegetable that should be detected (e.g., “carrot”). A variable number of images then followed, each displaying Gaussian noise applied to a homogenous white background, generated through MATLAB with the variance-level determined by the subject’s prior staircasing (see Fig. 1). In every run, each of the 4 fruit and vegetable cues was cumulatively followed by the same total amount of Gaussian noise. Within the blocks, the pure-noise images were each shown for 3 s. Following a variable length of time (between 12 and 24 s after the initial cue), a fruit or vegetable was presented, hidden within Gaussian noise using the parameters determined from the participant’s staircasing. The block automatically ended after this image. This design afforded us the high signal sensitivity found with block designs, combined with unpredictability to keep participants cognitively engaged. Each fruit and vegetable cue (e.g., “carrot”) occurred 3 times in a run (giving 12 blocks for each fruit and vegetable across the experiment). Blocks were presented in a pseudo-randomized order so the same cue did not immediately repeat. Two of the fruit-in-noise images ending the noise blocks contained a fruit or vegetable that did not match the preceding cue (i.e., 2 of 12 were foils) to focus participants on detecting the specific target. The 2 foils for each kind of cue were other fruits/vegetables with the same color but different shape, or same shape but different color (e.g., for carrot: tangerine and celery), ensuring that the 4 objects acted as foils with the same frequency. To encourage participants to search for the cued target from the very start of every block, the beginning of each run (during the 4 beginning time-points routinely removed in preprocessing) included a short block in which a cued fruit or vegetable appeared after only 3 or 6 s, followed by 12 s of fixation. The short blocks’ hidden fruits and vegetables were not repeated in the main blocks and the short blocks’ BOLD signal did not contribute to any analyses.

The 12 images of each type of fruit/vegetable (10 cued, 2 foils) that ended the blocks (within noise) were photographic examples in various orientations on a white background. The objects were all

adjusted to have the same height. The hidden items appeared in 4 possible locations: top-left, top-right, bottom-left, and bottom-right (see Fig. 1 for an example). The objects appeared in each location 3 times across the experiment, in a randomized order. The objects appearing in each location were preceded by the same cumulative amount of Gaussian noise across the experiment (i.e., there was no contingency between the amount of noise and final stimulus location).

Magnetic Resonance Imaging Preprocessing

Imaging data were preprocessed using the Analysis of Functional NeuroImages (AFNI) software package (Cox 1996). The first 4 volumes of each functional run were removed to allow the signal to reach steady-state magnetization. All functional images were slice-time corrected and a motion correction algorithm registered all volumes to a mean functional volume. Low-frequency trends were removed from all runs using a high-pass filter threshold of 0.01 Hz. Voxel activation was scaled to have a mean of 100, with a maximum limit of 200. The data were not smoothed.

Decoding Analyses

Pattern decoding was conducted within the MATLAB environment. The functional data were first *z*-scored within each run. The TRs containing neural activity in response to the pure-noise trials (i.e., before the participant encountered a concealed fruit or vegetable) were identified and then block-averaged in the following way. First, a binarized label of the search target was assigned to each TR. This time series of labels was then convolved with a time-shifted model of the hemodynamic response and thresholded at 0.8 (where the hemodynamic response falls between 0 and 1.0), giving a vector of activity values for each pure-noise TR. These vectors were then averaged by block. To ensure that each block's average was not influenced by the signal-ascent of the block's final trial (in which a fruit or vegetable was actually present), we also removed the last pure-noise TR of each block before averaging.

We conducted an information brain mapping "roaming searchlight" analysis in each participant by centering a sphere (3-voxel radius) on each voxel in turn (Kriegeskorte et al. 2006). MVPA was conducted with the voxels in each searchlight volume (123 when not restricted by the brain's boundary) and classifier performance was allocated to the central voxel. For each searchlight, 4-fold cross-validation was conducted (training on 3 runs; testing on the fourth) with a Gaussian Naïve Bayes (GNB) classifier (implemented through the MATLAB Statistics toolbox) to classify the activity to noise trials according to the participant's search target (carrot, celery, lime, or tangerine). The classifier was trained and tested on vectors of BOLD activity values that were identified in the manner described above. GNB classifiers have been shown to have particular success for datasets with small numbers of training samples (Ng and Jordan 2002; Mitchell et al. 2004; Singh et al. 2007), such as here, where each block contributes one data point. It is also fast for searchlight analyses (Pereira et al. 2009).

Each participant's map of searchlight accuracies was brought to standardized space (with the same resolution as the functional data) and spatially smoothed with a 6-mm FWHM kernel. The 11 searchlight maps were submitted to a group analysis to test whether the accuracy at each voxel was >0.25 (chance), with family-wise error correction for multiple comparisons (corrected to $P < 0.05$, with a 26-voxel cluster threshold estimated with AlphaSim; Cox 1996).

The next analysis tested whether a model trained on the cued visual noise would generalize to activity patterns (also block averages) recorded during the separate passive-viewing run. A classifier was trained on all pure-noise trials labeled by cue, and tested on the passive-viewing run of on-screen fruits and vegetables (after equalizing each pattern's mean through subtraction). This 4-way classification was performed with the voxels of each searchlight that had been identified in the prior analysis (transformed back into each participant's original space), with the searchlights' performance then averaged. We could not train on the passive-viewing data due to an insufficient amount of training data. To assess statistical significance, we conducted permutation testing. First, each participant's classifier testing labels were scrambled 1000 times, and the classification was repeated

for each new set of labels. This produced 1000 permutation-generated classification accuracies for each participant. To obtain a group *P*-value, a null distribution was created by randomly sampling a classification accuracy value from every subject's 1001 classification scores (1000 permutations + 1 real order) and calculating the group mean. This was performed 10 000 times, giving 10 000 permuted group means. The real group mean was compared with this null distribution to identify the *P*-value.

To conduct color and shape generalization tests, we trained classifiers to distinguish 2 items differing in one dimension (e.g., carrot vs. celery for color) and tested the model on the unused items that varied in the same way (e.g., tangerine vs. lime). This was performed on data from lateral occipital cortex, V4-region, and the left ATL region identified from the searchlight analysis. A 4-fold leave-one-run-out cross-validation procedure was conducted twice: Alternating which items were used for training. Each pair of scores was averaged. To assess statistical significance, we conducted the permutation testing procedure described above, with each set of randomized labels held constant for both combinations of training and testing. The null distribution was generated by sampling 1000 group means by randomly selecting from each participant's 100 permutations of classification scores. The *P*-value was then calculated from this distribution.

We conducted a novel decoding-dependency analysis (to investigate feature-to-identity convergence) by extracting classification accuracy vectors (i.e., 1 vs. 0 for each block) for color classification in the color region, shape classification in the shape region, and identity classification in the ATL searchlights. A logistic model (with quadratic penalty determined by marginal likelihood maximization for convergence [Zhao and Lyengar 2010] and coefficient stability) predicted object-identity decoding success (48 values; 1 for each block) for each of the identified ATL searchlights, with predicting variables for: block-by-block success for color decoding, block-by-block success for shape decoding and block-by-block color-shape conjunction (color decoding × shape decoding). Odds ratios were calculated from the models' coefficients (e^b) and averaged across the identified searchlights for each subject.

Regions of Interest

The color and shape across-item generalization tests were conducted using voxels in regions involved in shape and color processing. The shape-relevant region was based in lateral occipital cortex, an area with location-tolerant shape information (Eger et al. 2008; Carlson et al. 2011). Previous research has shown that this region is modulated by top-down processing (Stokes et al. 2009; Reddy et al. 2010). We extracted standard space coordinates from a highly cited study of shape processing (Grill-Spector et al. 1999). The lateral occipital shape region can be characterized by 3 vertices (dorsal, posterior, and anterior), so we placed 3 spheres (6 mm radius, the approximate standard deviation of activation reported) against the vertex coordinates from the object > texture contrast, in both hemispheres (coordinates in Table 1; Supplementary Fig. 1). This successfully encompassed lateral and ventral regions of the lateral occipital complex.

The color-processing region was based on a seminal color-processing study (McKeefry and Zeki 1997). The coordinates for maximum activation in a chromatic versus achromatic contrast were

Table 1
Coordinates for feature ROIs

Feature	Coordinates		
Shape	−41, −77, 3	−36, −71, −13	−38, −50, −17
	40, −72, 2	37, −69, −10	33, −47, −14
Color	30, −78, −18		
	−26, −80, −14		

Notes: Talairach coordinates for shape (extracted from Grill-Spector et al. 1999) and color (extracted from McKeefry and Zeki 1997) regions. The shape coordinates refer to the 3 vertices in each hemisphere that characterize the lateral occipital shape region. Spheres (6 mm radius) were positioned to border each vertex. The color coordinates reflect the center of right and left placed spheres (6 mm radius).

extracted from this study and a sphere (radius 6 mm, the approximate extent of activation reported) was placed at the right hemisphere coordinates. Investigations have suggested that right V4 is particularly modulated by top-down control of color processing (Kosslyn et al. 2000; Morita et al. 2004; Bramão et al. 2010), and achromatopsia is differentially associated with right V4 damage (Bouvier and Engel 2006), so we focused on the right region (coordinates in Table 1; Supplementary Fig. 1), although also examined left V4 from the same study.

Results

We presented participants with images of colored random noise and directed subjects to detect one of 4 types of fruits and vegetables—carrot, celery, lime, and tangerine—that vary systematically by shape and color (Fig. 1). These blocks of pure noise ended after an unpredictable amount of time with the cued fruit or vegetable, or a foil, hidden within noise. We analyzed the BOLD data associated with time-points before any fruit or vegetable was revealed, in order to examine top-down-driven activity. Prior to the main task, participants passively viewed exemplars of the 4 types of fruit and vegetable, giving us visually generated activity patterns for these items.

Decoding Object Identity from Anticipatory Visual Activity

To test the first requirement of a convergence zone—that a brain region contains a memory-evoked code for object identity—we asked if the identity (carrot, celery, lime, or tangerine) of a searched-for object could be decoded as participants viewed visual noise. The location (or even existence) of a convergence zone has not been established, so we used a whole-brain searchlight analysis to analyze sequential clusters of voxels. The functional data associated with the pure-noise time-points were labeled by the participant's current detection-target and submitted to a 4-way machine learning classifier. The classifier was able to decode (at $P < 0.05$ corrected) the identity of the anticipated-but-unseen targets in a cluster of 64 searchlights in the left ATL. The volume of the identified searchlights included the left fusiform gyrus, interior temporal, middle temporal, and superior temporal cortex (verified by cortical segmentation and automated labeling through FreeSurfer; Fischl et al. 2002). The region was centered at $-41x$, $-8y$, $17z$ and is shown in Figure 2. This was the only significant searchlight cluster (M accuracy = 0.29, $SD = 0.02$; confusion matrix available in Supplementary Fig. 2).

We verified that this significant decoding was not based purely on a subcategorical distinction between “fruits” and “vegetables” by successfully classifying items that do not cross this fruit/vegetable boundary (i.e., carrot vs. celery and lime vs. tangerine) at a level significantly above chance (permutation testing: $P = 0.025$). The “fruit versus vegetable” contrast itself was not classifiable in this region (M accuracy = 0.52; $P = 0.24$). We also confirmed that time-points from each of the 4 fruits and vegetables had above-chance accuracies ($P < 0.05$). Although unlikely that motor responses could account for temporal lobe performance, we confirmed that the numbers of motor responses did not differ significantly between targets ($F_{3,30} = 1.50$, $P = 0.24$).

We tested the specificity of the ATL's left lateralization by analyzing an ROI in the right hemisphere at the same y and z coordinates as the left region. Successful decoding was specific to the left ATL: the right ATL's performance was not significant (M accuracy = 0.26 where chance = 0.25; $P = 0.30$), with greater performance in the left ATL (paired t -test: $t_{(10)} = 3.64$, $P = 0.005$). Because of the known signal issues in the ATL, we measured the temporal signal-to-noise ratio (tSNR; calculated by dividing each voxel's mean signal with its standard deviation over the time-course of each run) of the left and right ATL regions, to assess signal-quality, and to ask if tSNR differences account for the lateralization. The tSNR values of the searchlight centers were high for both ATL regions (mean left = 77.4; mean right = 77.5) and well above levels that are considered suitable for signal detection (e.g., 20 in Binder et al. 2011). This indicated that the signal was strong in both regions, which additionally did not differ ($t_{(10)} = 0.01$, $P = 0.99$). Supplementary Figure 3 shows a map of ATL tSNR in this study.

Are multivoxel patterns necessary for distinguishing object identity? We would expect so, given the role of multivoxel patterns in successfully decoding object-information that cannot be detected with univariate analyses (Haxby et al. 2001; Eger et al. 2008). A direct and comparable approach to testing this is to re-run the classification, but replacing the multivoxel patterns with the univariate mean of each block (Coutanche 2013). Mean activation of the left ATL could not separate the conditions (M accuracy = 0.24; chance = 0.25; $P = 0.70$).

We examined the nature of the top-down-generated ATL identity code by asking whether it generalizes to activity recorded while subjects viewed images of the fruits and vegetables. We trained a classifier on the noise-only trials in the searchlights identified above (transformed back to each participant's original space), with each trial labeled by search target.



Figure 2. Location of searchlights with above-chance decoding of object identity while participants viewed visual noise and attempted to detect one of 4 kinds of fruit and vegetables. Left: A 4-way searchlight analysis revealed a region within the left ATL capable of decoding the target. Searchlight centers are shown in red. Right: The searchlights' volume displayed in one participant's original space, shown on their T_1 anatomical image after automated cortical reconstruction and volumetric segmentation using the FreeSurfer image analysis package (Fischl et al. 2002).

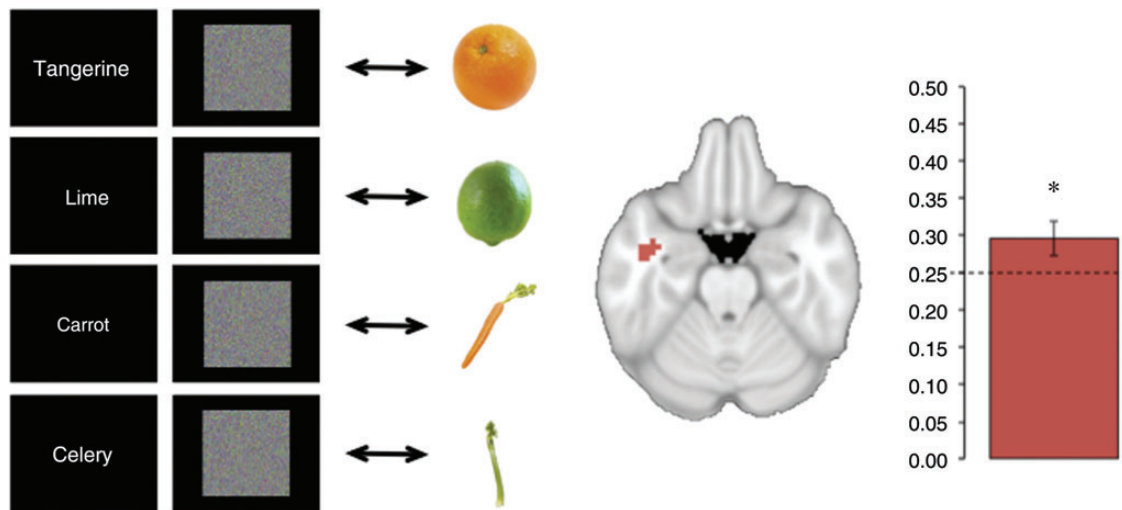


Figure 3. Generalizing from top-down activity to visual perception. Left: A classifier was trained on activity patterns recorded as participants viewed visual noise and sought to detect a cued fruit or vegetable. The classifier model was then tested on activity recorded as participants viewed real images of category examples in a separate run. Center: Activity patterns in this analysis were extracted from the left temporal lobe searchlights identified in the prior analysis of noise trials alone. Right: Classification accuracy significantly exceeded chance-performance, reflecting successful generalization from anticipatory activity to visual perception. The dashed line reflects the level of chance and the error bar shows the standard error of the mean. The asterisk signifies above-chance classification performance ($P < 0.05$).

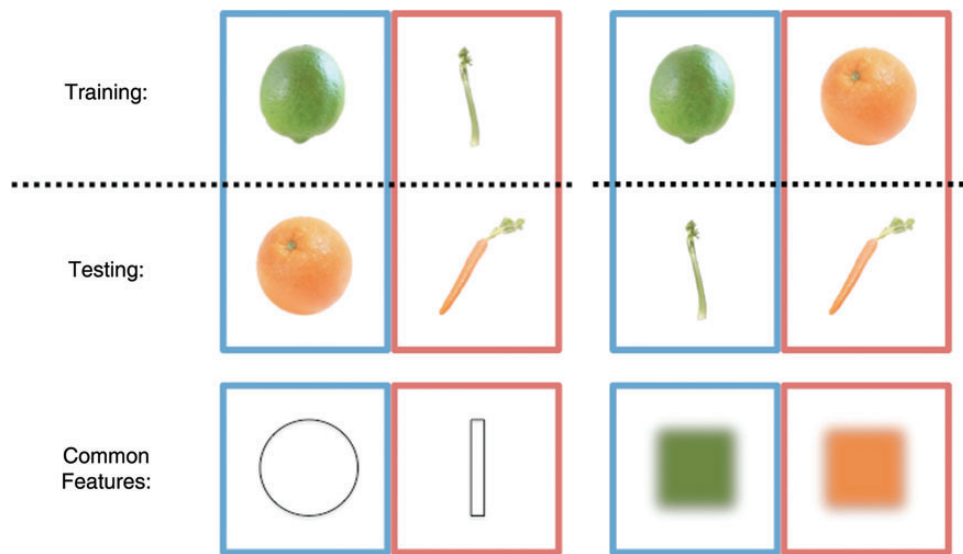


Figure 4. Feature-based generalization. Classifiers were trained to distinguish noise trials in which participants were searching for fruits and vegetables differing by shape or color. The classifiers were then tested on noise trials with the other pair of targets that differed in the same way. In the first example (left), classifiers are trained and tested based on shape (trained on lime vs. celery, tested on tangerine vs. carrot). In the second example (right), classifiers are trained and tested based on color (trained on lime vs. tangerine, tested on celery vs. carrot). The items took turns to act as the training data and the results of both comparisons were then averaged.

We tested the trained models on data from a separate run in which participants viewed blocked images of each kind of fruit or vegetable. The models trained on preparatory activity in the ATL could successfully classify the type of fruits and vegetable viewed in the separate passive-viewing run ($M = 0.30$, $SD = 0.08$; chance = 0.25; $P = 0.037$), revealing that the memory-generated and visually generated patterns were similarly structured (Fig. 3).

Decoding Object Features

The 4 targets in this study differed orthogonally by shape (2 elongated, 2 spherical) and color (2 orange, 2 green), allowing us to decode each feature independently, and test the second

prediction of the convergence zone theory: that specific feature knowledge fragments become active in sensory regions. We examined this by asking whether a model trained to distinguish different shapes or colors could generalize to another pair of objects that shares one dimension distinction, but differs in the other (Fig. 4). We investigated shape and color generalization in 1) a bilateral region of lateral occipital cortex that is associated with shape processing and 2) an occipital area (right V4) associated with color processing (see Materials and Methods for full details).

A classifier model trained on data collected as participants were searching for fruits and vegetables that differed by shape (e.g., lime vs. celery) could decode the remaining fruits and vegetables with a similar shape distinction (tangerine vs.

carrot), using activity from the bilateral lateral occipital cortex ($M=0.55$, $SD=0.06$; $P=0.01$; confusion matrix available in Supplementary Fig. 2). This ability to decode specific shapes across different colors (training on green, testing on orange) provides strong evidence that the specific features of each target were represented in the region's activity. The same region could not decode the color of the targets ($M=0.51$, $SD=0.07$; $P=0.43$), with higher performance for shape than for color ($P=0.05$). A right V4 region, in contrast, contained activity patterns that (using the same approach as above) could be used to decode color ($M=0.56$, $SD=0.08$; $P=0.01$; confusion matrix available in Supplementary Fig. 2), and shape at a trend level ($M=0.54$, $SD=0.09$; $P=0.09$; no significant difference between shape and color: $P=0.17$). This was specific to the right region: Left V4 could decode neither feature ($P>0.4$; lower color accuracies than the right: $P=0.04$, although no difference for shape: $P=0.15$). There was a significant interaction for greater shape decoding in lateral occipital cortex and greater color decoding in right V4; $P=0.03$; Fig. 5).

Unlike these features, identity decoding was unsuccessful in both regions ($P>0.46$). We also tested the identified ATL region for feature decoding. Consistent with this region containing identity information that is transformed away from features, neither shape ($M=0.49$, $SD=0.04$; $P=0.61$) nor color

($M=0.50$, $SD=0.05$; $P=0.46$) could be decoded (significantly lower color-decoding than right V4: $P=0.01$; lower shape-decoding than lateral occipital cortex: $P=0.01$; from permutation testing).

Shape and Color Conjunction Predicts the Left ATL's Identity Code

The third and final convergence zone prediction was that the convergence result (identity) would occur with converging activation of the specific shape and color feature fragments for that object. We employed a novel decoding-dependency analysis to examine this. We first coded each block of every participant for whether its neural activity contained decodable object identity in the ATL, color in right V4 and shape in lateral occipital cortex. We created a logistic regression model (full details in Materials and Methods) of identity-decoding success in the ATL (1 vs. 0) in each block, with predictors for the blocks' 1) color decoding success in right V4, 2) shape decoding success in lateral occipital cortex, and 3) simultaneous color-and-shape decoding (i.e., 1×2). The odds ratios from this model reflect dependencies between feature fragments and converged-upon identity.

The conjunction (i.e., convergence) of both V4 color decoding and lateral occipital shape decoding was specifically

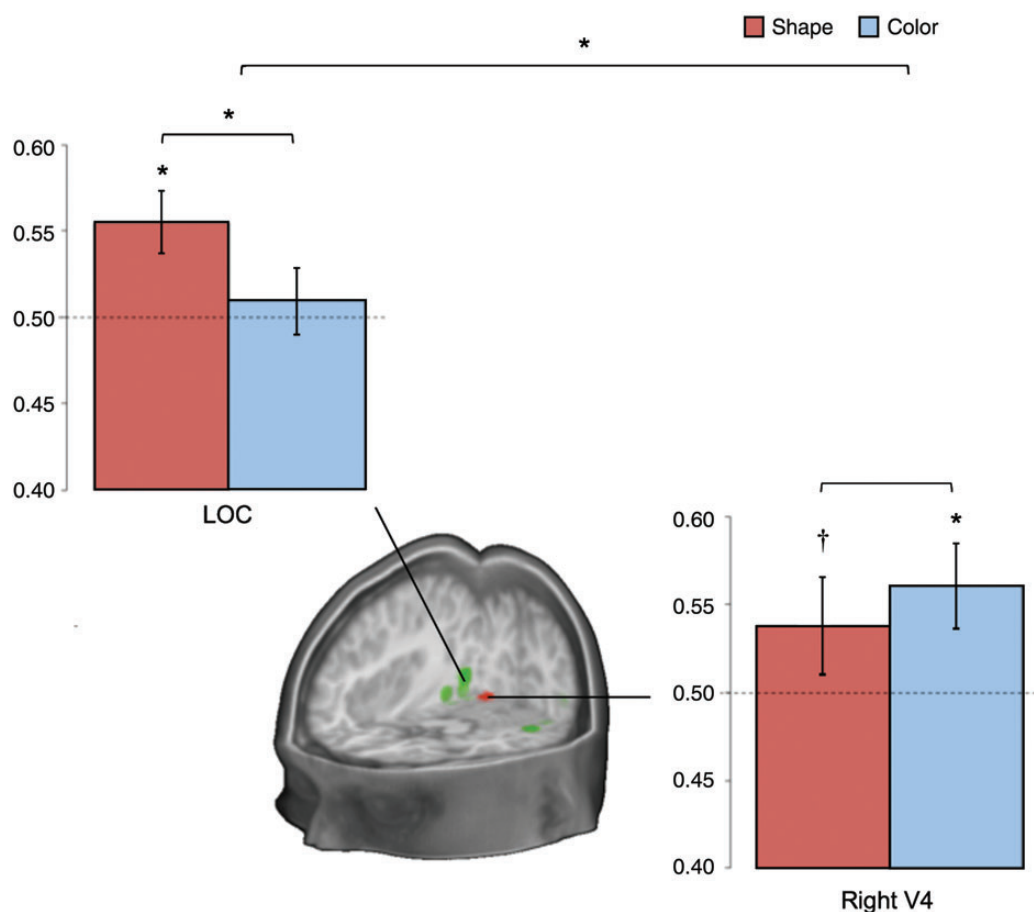


Figure 5. Classification results from the shape- and color-decoding analyses. Results are displayed from training a classifier on data from noise trials when participants were attempting to detect targets that differed by shape or color, and tested on data with other targets that varied in the same way. The shape results (e.g., training: lime vs. celery, testing: tangerine vs. carrot) are shown in red. The color results (e.g., training: tangerine vs. lime, testing: carrot vs. celery) are shown in blue. The dashed lines reflect the level of chance and the error bars show standard error of the mean. Asterisks signify above-chance classification performance ($P < 0.05$). The cross signifies trend-level performance ($P < 0.1$). The green region displayed in the cross section is in lateral occipital cortex. The red region is based on a color-responsive area, right V4 (Materials and Methods).

predictive of successful ATL identity decoding (M odds ratio = 2.64; odds ratio >1 : $t_{10} = 4.08$, $P = 0.002$), unlike the successful decoding of just one feature (M odds ratio for color: 0.76; M odds ratio for shape: 0.66). This relationship is also apparent from examining the amount of conjunctive color-and-shape decoding in ATL identity-decoded, and nondecoded, blocks. Randomly shuffling the color decoding and shape decoding results (10 000 times) provides a way to compare the observed degree of feature conjunction with a null distribution that controls for baseline rates of color and shape decoding success (as baseline rates remain the same in each permutation). Comparing the results to this null distribution showed that feature conjunction was more likely in the ATL identity-decoded blocks than expected under a null hypothesis of their independence ($P = 0.03$), but was not more likely in identity-misclassified blocks ($P = 0.38$; interaction: $P = 0.07$). This shows that concurrent color-and-shape decoding (in feature regions) co-occurs with successful object-identity decoding (in the ATL) to a greater degree than expected from baseline occurrences. The overall probability of simultaneous color-and-shape decoding was not significantly greater than expected from independence ($t = -1.28$, $P = 0.23$), further emphasizing the relevance of successful ATL decoding.

The above dependency relationship was not driven by accuracy or reaction times during the task: Blocks with, and without, synchrony between color, shape, and identity decoding did not significantly differ in behavioral accuracy ($t_{(10)} = 0.57$, $P = 0.58$) or (correct) response times ($t_{(10)} = 0.76$, $P = 0.46$). To examine if convergence requires feature information coming from independent regions, or if distinct features can also be extracted from one region, we also ran a model with shape decoding from V4 rather than lateral occipital cortex. In this model, shape-color conjunction was also the only factor receiving an odds ratio that was greater than one ($t_{(10)} = 2.58$, $P = 0.03$).

Finally, if the relationship between color-shape conjunction and ATL identity decoding plays an active role in the activation of a concept, we might expect subjects with stronger dependencies to have ATL codes that more closely match visually generated codes. Consistent with this, subjects with stronger links between feature fragments and identity (indicated by higher odds ratios for shape-and-color convergence predicting ATL identity in the previous logistic regression) had top-down identity codes that more closely resembled visually driven patterns in the ATL ($r = 0.67$, $P = 0.02$; Fig. 6). This was not simply due to individual differences in the robustness of top-down-driven activity: the strength of the feature-identity dependency was not related to decoding success of a classifier trained and tested on noise only ($P = 0.72$), suggesting that the relationship was specific to similarity between memory-generated and visually generated codes.

Discussion

The results described here provide evidence for theories that posit a region of integration (Damasio 1989; Simmons and Barsalou 2003; Patterson et al. 2007; Meyer and Damasio 2009). As participants viewed visual noise and attempted to detect a specific fruit or vegetable, activity patterns of the left ATL coded the retrieved object's identity. These top-down patterns generalized to activity produced during passive viewing. Posterior featural regions encoded the anticipated objects' specific

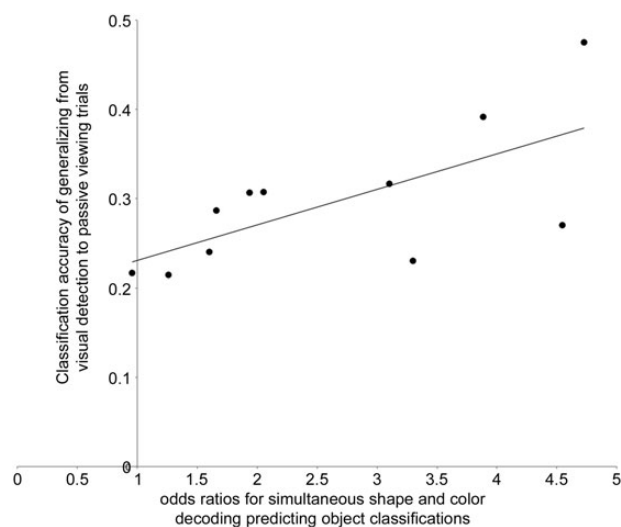


Figure 6. Individual differences in noise-to-visual generalization against the strength of the relationship between featural- and object-identity decoding. The y-axis represents each subject's classification performance from training on cued noise and testing on visual presentations of each fruit and vegetable in the ATL. The x-axis reflects each participant's odds ratio for the conjunction of color and shape decoding (in relevant feature regions) predicting cued-noise identity classifications in the ATL. A logistic regression model generated the odd ratios (details in Materials and Methods).

shape and color. Importantly, these levels of representation were closely linked: ATL decoding of object identity was predicted by the presence of both color and shape codes (but neither one alone) in featural regions. Further, a stronger feature-to-identity link predicted a greater match between top-down and visually generated patterns in the ATL.

A recent MVPA study reported that the ATL contains information relating to conceptual dimensions (object location and action) of visually presented objects, while occipitotemporal cortex does not (Peelen and Caramazza 2012). Here, we found that the ATL's code for an object's identity is linked to occipitotemporal codes for the object's distinct visual properties. Additionally, by examining activity during a top-down task (rather than during visual presentations), we found that these distinct, but connected, codes do not require visual stimulation.

The particular ATL site we identified directly overlaps with reported peaks of the semantic network (Vandenberghe et al. 1996), and has been associated with semantic behavioral disruptions when targeted with TMS (e.g., Pobric et al. 2010a, 2010b; Ishibashi et al. 2011). Interestingly, this region was also recently identified in a study of conceptual combination (Baron and Osherson 2011). Here, the authors found that multivoxel patterns in this cortical location encode the semantic relationship between word combinations (such as "man" and "child") and related concepts (boy).

The results of our study are consistent with a variety of patient work pointing to the ATL's role in conceptual knowledge (Patterson et al. 2007; Rogers et al. 2007). More specifically, the reported translation from featural information within visual cortex to identity information in the ATL is consistent with the temporal lobes containing a continuous transition to higher order representations (Binney et al. 2012). It should be noted that information may still be present within additional regions of the ATL. The ventral ATL is particularly vulnerable to signal dropout and distortion, potentially obscuring an informative

signal in this area (although recent distortion-correcting procedures have identified this region as active during semantic tasks: Visser, Embleton et al. 2010; Visser and Lambon Ralph 2011; Visser, Jefferies et al. 2012). As such, information may also lie in neighboring ATL regions that did not have a robust signal.

The information-dependency we report between ATL and visual cortex identifies a top-down-generated relationship between these areas, where visual regions evoke specific feature activity patterns that are functionally related to ATL patterns. As shape and color processing are known to be neurally dissociable (e.g., agnosia patients can show impairments in one but not the other: Cavina-Pratesi et al. 2010), this suggests that distinct systems are sharing information with the ATL and contributing different types of conceptual information (Lambon Ralph et al. 2010). The hub-and-spoke model proposes that modality-specialized regions (V4 and lateral occipital cortex in this study) provide sensory and motor substrates that are combined into an independent high-dimensional representational space in a central hub (Lambon Ralph et al. 2010; Pobric et al. 2010a). Our finding of a predictive link between feature and identity coding regions gives weight to sensory regions playing a significant role in object knowledge. The distinct roles of the ATL and sensory regions are consistent with reports that ATL atrophy can increase reliance on superficial similarities (coded in sensory regions) rather than semantic similarities (coded in the ATL; Lambon Ralph et al. 2010). An account with multiple convergence zones (Damasio 1989) also incorporates distinct roles for integration and sensorimotor regions. Further investigations are needed to distinguish between these theories.

The findings of this study may be helpful in interpreting some recent semantic dementia patient findings. A recent study of semantic dementia reported that processing items rich in visual color and form was disproportionately impaired in patients with severe dementia, unlike items with other features such as sound/motion and tactile/action (Hoffman et al. 2012). The authors speculated that temporal lobe atrophy may have spread more posteriorly to affect basic featural regions in these severe cases. Our results suggest a new possibility: the patients may have experienced disruption to a key shape and color convergence zone. One arising question for the hub-and-spoke model (which includes a transmodal hub) is why visual properties would be affected, while other sensory inputs are not. One possibility is that parts of the hub's neural network that are closer to posterior occipitotemporal cortex are particularly important for concepts with a strong dependency on these visual features. Probabilistic tractography recently revealed connection heterogeneity across the ATL (Binney et al. 2012). Distinctions may also be present for concepts that particularly depend on one modality more than others.

Our between-subject analyses showed that participants differed in the correspondence between top-down and visually generated ATL patterns, which was predicted by the strength of their ATL's dependence on joint color-and-shape decoding. One of several explanations could underlie this result. One possibility is that top-down generated patterns only take on perceptually driven characteristics when all relevant early visual regions generate robust and specific features, which in turn varies across blocks and participants. Alternatively, individual differences in whether a person's top-down patterns rely on sensory features (or more abstract properties; cf. Hsu et al. 2011) could determine whether top-down processes activate all relevant sensory information. Our individual difference result

raises several fascinating questions for future research: Do differences in the link between object identity and feature synchrony produce differences in people's phenomenological experiences during processes involving retrieval, such as imagery? Are time-points with synchronous color-and-shape decoding accompanied by particularly vivid imagery?

There are a number of reasons to be confident that the object-identity decoding reported here reflects visual processes, rather than others, such as verbal rehearsal. The link between object-identity and shape-and-color decoding argues strongly for a perceptual basis for the identity decoding, rather than other semantic features such as taste. Further, the ability to generalize decoding from top-down to visually presented objects (where no task was required) supports a visual account.

In our featural analyses, we found shape, but not color, information in the shape region. In contrast, the color region had decodable color information along with shape information at a trend level. Interestingly, this asymmetry was also reported in a recent meta-analysis of modality-specific imagery, where shape-related activity overlapped with color regions, but not vice versa (McNorgan 2012). Prior work has also suggested that "V4 neurons are at least as selective for shape as they are for color" (Roe et al. 2012, p. 17). Shape curvature is particularly represented in V4 (Roe et al. 2012), which would account for V4 decoding spherical versus elongated shapes at a trend here. We note that our interpretation of the convergence pattern does not require that the two regions are uniquely selective to color or shape; only that they contain different (i.e., nonredundant) patterns of information.

Several neural and cognitive states might underlie the successful decoding of just one of the two features during a block. On the one hand, fluctuations in a specific feature might reflect block-by-block changes in participants' cognitive strategies, such as greater attention directed to one feature. On the other hand, fluctuations in neural activity might reflect ongoing competition between the features for limited attentional resources. When the attentional system becomes taxed, dynamic competition between features may lead to only one becoming fully activated and having discriminable activity patterns.

We employed a novel decoding-dependency analysis in this work, enabling us to identify a link between ATL's object-identity code and a conjunction of visual feature decoding in occipital regions. This type of analysis has great potential for future investigations of other configural stimuli, such as in multisensory interplay (Driver and Noesselt 2008), to test whether the synchronous emergence of composing features co-occurs with the generation of a higher level code. Relating this measure to between-subject differences, as we have done here, can help elucidate the behavioral and neural consequences of such links. Employing four fruits and vegetables here in a 2-by-2 design allowed us to independently evaluate shape, color and identity in a systematic manner, but we acknowledge that the actual semantic space for known objects is vast. Future studies may wish to further explore the semantic relations between large numbers of items.

In summary, this study has found that the top-down retrieval of object knowledge leads to activation of shape-specific and color-specific codes in relevant specialized visual areas, as well as an object-identity code within left ATL. Moreover, the presence of identity information in left ATL was more likely when both shape and color information was simultaneously present in respective feature regions. The strength of this dependency

predicted the correspondence between top-down and bottom-up activity patterns in the ATL. These findings support proposals that the ATL integrates featural information into a less feature-dependent representation of identity.

Supplementary Material

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

Funding

This work was funded by National Institutes of Health grants R01MH070850 and R01EY021717 (to S.L.T.-S.). M.N.C. was funded by a fellowship from the Howard Hughes Medical Institute.

Notes

We thank C.A. Gianessi and P. van den Berg for assistance with stimulus development, and R.A. Epstein, J.W. Kable, L.S. Hallion, M.A. Lambon Ralph, T.T. Rogers, and an anonymous reviewer for their insightful comments. *Conflict of Interest*: None declared.

References

- Baron SG, Osherson D. 2011. Evidence for conceptual combination in the left anterior temporal lobe. *Neuroimage*. 55:1847–1852.
- Binder JR, Gross WL, Allendorfer JB, Bonilha L, Chapin J, Edwards JC, Weaver KE. 2011. Mapping anterior temporal lobe language areas with fMRI: A multicenter normative study. *Neuroimage*. 54(2): 1465–1475.
- Binney RJ, Parker GJM, Lambon Ralph MA. 2012. Convergent connectivity and graded specialization in the rostral human temporal lobe as revealed by diffusion-weighted imaging probabilistic tractography. *J Cogn Neurosci*. 24(10):1998–2014.
- Bouvier SE, Engel SA. 2006. Behavioral deficits and cortical damage loci in cerebral achromatopsia. *Cereb Cortex*. 16(2):183–191.
- Bozeat S, Lambon Ralph MA, Patterson K, Garrard P, Hodges JR. 2000. Non-verbal semantic impairment in semantic dementia. *Neuropsychologia*. 38(9):1207–1215.
- Bramão I, Faísca L, Forkstam C, Reis A, Petersson KM. 2010. Cortical brain regions associated with color processing: an fMRI study. *Open Neuroimaging J*. 4:164–173.
- Carlson T, Hogendoorn H, Fonteijn H, Verstraten FAJ. 2011. Spatial coding and invariance in object-selective cortex. *Cortex*. 47(1):14–22.
- Cavina-Pratesi C, Kentridge RW, Heywood CA, Milner AD. 2010. Separate processing of texture and form in the ventral stream: evidence from FMRI and visual agnosia. *Cereb Cortex*. 20(2):433–446.
- Coutanche MN. 2013. Distinguishing multi-voxel patterns and mean activation: why, how, and what does it tell us? *Cogn Affect Behav Neurosci*. 13(3):667–673.
- Cox RW. 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 29(3):162–173.
- Damasio AR. 1989. Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition*. 33(1–2):25–62.
- Driver J, Noesselt T. 2008. Multisensory interplay reveals crossmodal influences on “sensory-specific” brain regions, neural responses, and judgments. *Neuron*. 57(1):11–23.
- Eger E, Ashburner J, Haynes J-D, Dolan RJ, Rees G. 2008. fMRI activity patterns in human LOC carry information about object exemplars within category. *J Cogn Neurosci*. 20(2):356–370.
- Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S et al. 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 33(3):341–355.
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R. 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*. 24(1):187–203.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*. 293(5539):2425–2430.
- Hodges JR, Patterson K, Oxbury S, Funnell E. 1992. Semantic dementia progressive fluent aphasia with temporal lobe atrophy. *Brain*. 115(6):1783–1806.
- Hoffman P, Jones RW, Lambon Ralph MA. 2012. The degraded concept representation system in semantic dementia: damage to pan-modal hub, then visual spoke. *Brain*. 135(12):3770–3780.
- Hsu NS, Kraemer DJM, Oliver RT, Schlichting ML, Thompson-Schill SL. 2011. Color, context, and cognitive style: variations in color knowledge retrieval as a function of task and subject variables. *J Cogn Neurosci*. 23(9):2544–2557.
- Ishibashi R, Lambon Ralph MA, Saito S, Pobric G. 2011. Different roles of lateral anterior temporal lobe and inferior parietal lobule in coding function and manipulation tool knowledge: evidence from an rTMS study. *Neuropsychologia*. 49(5):1128–1135.
- Kiefer M, Pulvermüller F. 2012. Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex*. 48(7):805–825.
- Kosslyn SM, Thompson WL, Costantini-Ferrando MF, Alpert NM, Spiegel D. 2000. Hypnotic visual illusion alters color processing in the brain. *Am J Psychiatry*. 157(8):1279–1284.
- Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *Proc Natl Acad Sci USA*. 103(10):3863–3868.
- Lambon Ralph MA. 2014. Neurocognitive insights on conceptual knowledge and its breakdown. *Philos Trans R Soc B Biol Sci*. 369(1634):20120392.
- Lambon Ralph MA, Sage K, Jones RW, Mayberry EJ. 2010. Coherent concepts are computed in the anterior temporal lobes. *Proc Natl Acad Sci USA*. 107(6):2717–2722.
- Luzzi S, Snowden JS, Neary D, Coccia M, Provinciali L, Lambon Ralph MA. 2007. Distinct patterns of olfactory impairment in Alzheimer's disease, semantic dementia, frontotemporal dementia, and corticobasal degeneration. *Neuropsychologia*. 45(8):1823–1831.
- Martin A. 2007. The representation of object concepts in the brain. *Annu Rev Psychol*. 58:25–45.
- McKeefry DJ, Zeki S. 1997. The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain*. 120(12):2229–2242.
- McNorgan C. 2012. A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. *Front Hum Neurosci*. 6:285.
- Meyer K, Damasio A. 2009. Convergence and divergence in a neural architecture for recognition and memory. *Trends Neurosci*. 32(7):376–382.
- Mitchell T, Hutchinson R, Niculescu R, Pereira F, Wang X, Just M, Newman S. 2004. Learning to decode cognitive states from brain images. *Machine Learn*. 57(1):145–175.
- Morita T, Kochiyama T, Okada T, Yonekura Y, Matsumura M, Sadato N. 2004. The neural substrates of conscious color perception demonstrated using fMRI. *Neuroimage*. 21(4):1665–1673.
- Ng AY, Jordan MI. 2002. On discriminative vs. generative classifiers: a comparison of logistic regression and naive bayes. *Adv Neural Inf Process Syst*. 2(14):841–848.
- Patterson K, Nestor PJ, Rogers TT. 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci*. 8(12):976–987.
- Peelen MV, Caramazza A. 2012. Conceptual object representations in human anterior temporal cortex. *J Neurosci*. 32(45):15728–15736.
- Pereira F, Mitchell T, Botvinick M. 2009. Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*. 45(1 Suppl): S199–S209.
- Pobric G, Jefferies E, Lambon Ralph MA. 2010b. Amodal semantic representations depend on both anterior temporal lobes: evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia*. 48(5):1336–1342.

- Pobric G, Jefferies E, Lambon Ralph MA. 2007. Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. *Proc Natl Acad Sci USA*. 104 (50):20137–20141.
- Pobric G, Jefferies E, Lambon Ralph MA. 2010a. Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Curr Biol*. 20(10):964–968.
- Reddy L, Tsuchiya N, Serre T. 2010. Reading the mind's eye: decoding category information during mental imagery. *Neuroimage*. 50 (2):818–825.
- Roe AW, Chelazzi L, Connor CE, Conway BR, Fujita I, Gallant JL, Lu H, Vanduffel W. 2012. Toward a unified theory of visual area V4. *Neuron*. 74(1):12–29.
- Rogers TT, Lambon Ralph MA, Garrard P, Bozeat S, McClelland JL, Hodges JR, Patterson K. 2004. Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychol Rev*. 111(1):205–235.
- Rogers TT, Patterson K, Graham K. 2007. Colour knowledge in semantic dementia: it is not all black and white. *Neuropsychologia*. 45 (14):3285–3298.
- Simmons WK, Barsalou LW. 2003. The similarity-in-topography principle: reconciling theories of conceptual deficits. *Cogn Neuropsychol*. 20(3):451–486.
- Singh V, Miyapuram KP, Bapi RS. 2007. Detection of cognitive states from fMRI data using machine learning techniques. *Proceedings of Twentieth International Conference on Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann Publishers Inc. p. 587–592.
- Stokes M, Thompson R, Nobre AC, Duncan J. 2009. Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proc Natl Acad Sci USA*. 106(46):19569–19574.
- Vandenberghe R, Price C, Wise R, Josephs O, Frackowiak RSJ. 1996. Functional anatomy of a common semantic system for words and pictures. *Nature*. 383:254–256.
- Visser M, Embleton KV, Jefferies E, Parker GJ, Lambon Ralph MA. 2010. The inferior, anterior temporal lobes and semantic memory clarified: novel evidence from distortion-corrected fMRI. *Neuropsychologia*. 48(6):1689–1696.
- Visser M, Jefferies E, Lambon Ralph MA. 2010. Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *J Cogn Neurosci*. 22(6):1083–1094.
- Visser M, Jefferies E, Embleton KV, Lambon Ralph MA. 2012. Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *J Cogn Neurosci*. 24(8):1766–1778.
- Visser M, Lambon Ralph MA. 2011. Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. *J Cogn Neurosci*. 23(10):3121–3131.
- Zhao M, Lyengar S. 2010. Nonconvergence in logistic and poisson models for neural spiking. *Neural Comput*. 22(5):1231–1244.