Published in final edited form as:
J Cogn Neurosci. 2014 December ; 26(12): 2735–2749. doi:10.1162/jocn_a_00661.

Feature diagnosticity affects representations of novel and familiar objects

Nina S. Hsu[1,2], **Margaret L. Schlichting**[1,3], and **Sharon L. Thompson-Schill**[1]

[1]Department of Psychology, Center for Cognitive Neuroscience, University of Pennsylvania

[2]Department of Psychology, Center for Advanced Study of Language, University of Maryland, College Park

[3]Department of Psychology, University of Texas at Austin

Abstract

Many features can *describe* a concept, but only some features *define* a concept in that they enable discrimination of items that are instances of a concept from (similar) items that are not. We refer to this property of some features as feature diagnosticity. Previous work has described the behavioral effects of feature diagnosticity, but there has been little work on explaining why and how these effects arise. In this study, we aimed to understand the impact of feature diagnosticity on concept representations across two complementary experiments. In Experiment 1, we manipulated the diagnosticity of one feature, color, for a set of novel objects that human subjects learned over the course of one week. We report behavioral and neural evidence that diagnostic features are likely to be automatically recruited during remembering. Specifically, individuals activated color-selective regions of ventral temporal cortex (specifically, left fusiform gyrus and left inferior temporal gyrus) when thinking about the novel objects, even though color information was never explicitly probed during the task. Moreover, multiple behavioral and neural measures of the effects of feature diagnosticity were correlated across subjects. In Experiment 2, we examined relative color association in familiar object categories, which varied in feature diagnosticity (fruits and vegetables, household items). Taken together, these results offer novel insights into the neural mechanisms underlying concept representations by demonstrating that automatic recruitment of diagnostic information gives rise to behavioral effects of feature diagnosticity.

Introduction

Any concept, such as a lion, can be described by a list of properties or features, and these features will vary in terms of how common they are among concepts (e.g., *alive*), how unique they are (e.g., *king-of-the-jungle*), how strongly associated they are with the concept (e.g., *loud-roar*), how behaviorally-relevant they are (e.g., *attacks-humans*), and so on. For any given pair of concepts (e.g., lion and tiger), some features will be diagnostic for

Address correspondence to: Nina S. Hsu, Center for Advanced Study of Language, 7005 52nd Avenue, University of Maryland, College Park, College Park, MD 20742 USA, Phone: (301) 226-9150, ninahsu@umd.edu.

Conflict of interest:
The authors declare no competing financial interests.

distinguishing between them (e.g., *has-stripes*) and others will not (e.g. *has-fur*). The goal of these studies is to understand the impact that these sorts of variables have on the representation of concepts, and our specific focus is on the notion of *feature diagnosticity*. The diagnostic feature in question is color, motivated in part by the growing literature showing that visual brain systems are recruited when thinking about concepts that have a specific visual feature (e.g., recruiting color-sensitive brain areas when subjects remember colorful concepts like fruits) (Chao & Martin, 1999; Hsu, Frankland, & Thompson-Schill, 2012; Hsu, Kraemer, Oliver, Schlichting, & Thompson-Schill, 2011; Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995; Simmons et al., 2007).

There are a number of studies that *describe* the effects of diagnosticity on behavior; however, we do not believe that there currently exists a mechanism to explain how or why these effects arise. For example, although participants can perceive diagnostic features of an object as easily as non-diagnostic features, they selectively attend to those features which are most useful for discrimination (Schyns, 1998). Subjects name objects with highly diagnostic colors faster and with fewer errors than for objects with non-diagnostic colors (Tanaka & Presnell, 1999), while children can be trained to attend to object shape in the context of naming, leading to faster object naming times (Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002). Further, feature verification tasks have shown that diagnostic features hold a privileged status in an object's overall representation, as subjects' responses were faster when the feature was diagnostic of the concept than when the feature was shared amongst other category members (Cree, McNorgan, & McRae, 2006). We find these results intriguing, but lacking in providing a mechanism as to why feature diagnosticity affects behavior the way it does.

Similarly, there are a handful of neurophysiological findings that examine the impact of feature diagnosticity on neural measures. Single-unit and local field potential studies have shown selective tuning of neurons in response to relevant features. In macaque monkeys, inferotemporal (IT) neurons showed an increased response to diagnostic features, depending on the importance of those features for object categorization (Sigala & Logothetis, 2002). Neurons in the anterior IT cortex also responded similarly to images showing either 10% or 50% relevant information (Nielsen, Logothetis, & Rainer, 2006). This region-specific insensitivity to the stimulus image itself was coupled with a graded response to behaviorally relevant features in the posterior IT cortex. Thus, stimulus features can be preferentially represented if they are diagnostic for a behavior, and the neural representation of an object can be influenced by both visual experience and viewing history.

These studies provide descriptions rather than explanations of diagnosticity effects; in part, these effects are difficult to understand because so many variables are confounded in conceptual structure. In order to measure the impact of a single variable – feature diagnosticity – on concept representations, we created and taught subjects a set of novel objects. In this way, we could control the structure of the conceptual space and thereby eliminate those confounds that are unavoidable with real world objects (Grossman, Blake, & Kim, 2004; James & Gauthier, 2003; Kiefer, Sim, Liebich, Hauk, & Tanaka, 2007; Weisberg, van Turennout, & Martin, 2007). For example, "barks" is a diagnostic feature of

dogs, but it is also an uncommon feature in the animal kingdom; the object concepts in our artificial world have features varying in diagnosticity while holding frequency constant.

The experiments described here employ both univariate and multivariate techniques in order to measure the impact of feature diagnosticity on concept representations. Recent neuroimaging studies utilizing multivariate methods have demonstrated that patterns of brain activation, as opposed to averaged overall regional activation, can carry meaningful information (Cox & Savoy, 2003; Haxby et al., 2001; Kamitani & Tong, 2005). Multivariate analyses have been used to decode categories of remembered stimuli (Polyn, Natu, Cohen, & Norman, 2005), compare similarity of disparate categories (O'Toole, Jiang, Abdi, & Haxby, 2005), and decode neural similarity within a single object category of abstract shape (H. P. Op de Beeck, Torfs, & Wagemans, 2008), or a single natural category of mammals (Weber, Thompson-Schill, Osherson, Haxby, & Parsons, 2009). These multivariate analyses add a complementary approach to the extant fMRI literature (Jimura & Poldrack, 2012).

In Experiment 1, subjects learned a set of 12 novel objects. Half of the subjects learned that the conjunction of color and shape was diagnostic of object category (henceforth referred to as the CS group); the other half of the subjects learned that shape was sufficient to distinguish amongst the set of objects (the S group). Critically, we matched color variability amongst both sets (i.e., two objects were purple, two objects were green, etc.). Following training, we collected a variety of behavioral measures, and we measured fMRI responses during a test for memory of the shape of the objects. We hypothesized that accessing diagnostic feature information would result in group differences behaviorally and neurally; specifically, we would observe group differences in the activation of color-selective brain regions. Both behavioral and neural measures revealed effects of our manipulation of feature diagnosticity. Subjects in both groups learned the colors of the objects equally well; however, compared to S subjects, CS subjects more frequently used color to describe the objects. Moreover, they activated ventral temporal cortex (specifically, left fusiform gyrus and left inferior temporal gyrus) even when color information was not explicitly probed. Further, a multivariate measure of neural similarity predicted color similarity ratings for the CS group only. Experiment 2 examined relative color association in familiar object categories: fruits and vegetables (FV) and household items (HHI). In addition to providing a useful test of the generalization of the Experiment 1 results to a separate set of categories, Experiment 2 replicated many results from Experiment 1 and provided some interesting contrasts. Together, our results suggest that diagnostic features are more likely to be accessed automatically than are non-diagnostic features during remembering, and that automatic recruitment of diagnostic information gives rise to the behavioral effects of feature diagnosticity. The features that we use to categorize objects and not simply the features that we explicitly remember about objects shape the neural representations of object concepts.

Materials and Methods

Experiment 1

Participants—Sixty-three (n = 63) healthy subjects participated in the study (17 males, 46 females; average age = 22.8 years, range = 18–30 years). Thirty-two (n = 32) of these

subjects participated in the subsequent fMRI portion of the study (9 males and 23 females; average age = 24.7 years, range = 18–30 years). All subjects were right-handed, native speakers of English, and were not taking any psychoactive medications over the course of the study. Those subjects participating in the fMRI portion had no history of neurological disorders and a healthy neurological profile. We paid subjects $10/hour for behavioral portions of training, and $20/hour for participating in the fMRI portion. Subjects provided written informed consent to participate, and the human subjects review board at the University of Pennsylvania approved all experimental procedures.

Training Materials and Procedure—In a between-subjects design, subjects were randomly assigned to learn one of two object sets. In the “color+shape” (CS) set, color is necessary for object identification, and shape information is not sufficient (e.g., objects have similar shapes but differ in color, like lemons and limes). In the “shape” (S) set, color is available for object identification, but shape information is sufficient (e.g., objects differ in both shape and color, like stop signs and yield signs). Subjects learned one of these objects sets over the course of four 30–60 minute training sessions that took place over seven days.

Stimuli: For either object set, as shown in Figure 1, subjects were trained on a set of 36 exemplars of 12 distinct object basic level categories (3 exemplars per category). Each category had a pseudoword name (Rastle, Harrington, & Coltheart, 2002). Stimuli were created from scratch in Blender 2.48 (www.blender.org). All objects were given the same surface texture and illuminated with the same single light source. For object shape, we created four shape variants for the CS set, and twelve shape variants for the S set (four of these shape variants overlapped with those in the CS set: *fulch, hinch, klarve, screll*). For object size, we created two additional size variants for each exemplar by halving or doubling the scale of the object, thus creating three possible sizes for each object. Size was an irrelevant dimension for object identification, but did serve to distinguish the individual exemplars within each basic category. For object color, we used Blender’s HSV color space, in which color is determined from a set of three values, one each corresponding to hue, saturation and value (or luminance). We held saturation and value constant, varying hue in order to create six distinct color categories for the objects.

The differences in the sets are summarized in Figure 1. Two properties are critical between the two sets. First, note that in order to identify successfully each object by distinguishing it from the others in the set (i.e., when $P(\text{object}) = 1.00$), the CS set requires the conjunction of shape and color information, whereas the S set only requires shape information. Second, color probability (i.e., that $P(\text{object} \mid \text{color}) = 0.50$) is matched between the two groups. Thus, the two groups differ in terms of fixed diagnosticity (with color for CS objects being *relatively* more diagnostic than color for S objects).

For each of the 36 exemplars, we created 10-second videos of each exemplar rotating 360 degrees, counterclockwise, on a raised, black platform against a gray background. For all behavioral tasks described below, we used PsyScope (http://psy.cns.sissa.it) to present stimuli and collect responses and response times (RT). The training schedule and list of tasks can be found in Table 1.

Training, Video Exposure: Subjects viewed a randomized sequence of videos, with each video individually presented. Each video was shown twice. This sequential presentation resulted in 72 videos that played for approximately 12 minutes. While each video played, the exemplar name appeared below, and subjects were instructed to repeat aloud the name of the object currently being viewed. Subjects watched the videos at the beginning of the first, second, and third sessions.

Training, Naming: We assessed knowledge of the novel objects through a naming task. Immediately after viewing the object videos, subjects saw a screenshot of each of the 36 exemplars. Upon typing that exemplar's name, they were given corrective feedback. Subjects participated in the naming task during each training session, viewing a total of 12 trials of each object category across all four sessions. We used four unique screenshots for each exemplar (at the 50th, 100th, 150th, and 200th frame of the video, counterbalanced across the size variants for each object), such that subjects never viewed identical images for the same exemplar (or for the same object) during this task. Three subjects who did not exceed 80% accuracy on recalling object names (measured via the naming task) during the fourth session were excluded from further analysis.

Testing, Adjective Generation Task: Feature listing has previously been used as a measure of diagnosticity (Tanaka & Presnell, 1999), where features considered to be diagnostic are listed earlier and more often than other features. For each trial of this task, we presented subjects with an object name. They were instructed to list 2–4 adjectives describing the object. Subjects could proceed at their own pace by pressing the ENTER key to proceed to the next trial. We administered this task last during the third session.

Testing, Pairwise General Similarity Task: We assessed psychological similarity by having subjects rate the general similarity of every pairing of the 12 learned novel objects, resulting in 66 pairwise ratings. For each trial of this task, we presented subjects with a pair of object names, along with a scale numbered from 1 (very dissimilar) to 9 (very similar). Subjects assigned each similarity rating at their own pace, and each response triggered the beginning of the next pairing. We administered this task during the fourth session, after the naming task but before the color naming task (see below). Note that subjects were told to base their judgments on general similarity, and they were not asked to base their ratings on any particular object features.

Testing, Color Naming Task: At the end of the fourth session, the experimenter verbally named each of the objects individually. Subjects were instructed to report the color that they associated with that object, and the experimenter recorded the response.

Untrained Similarity Rating: In order to assess the relationship between behavioral similarity and neural similarity, we obtained psychological similarity ratings based on perception (i.e., pictures) rather than memory (i.e., names) of the novel objects. Thus, we slightly modified the pairwise similarity task described earlier by having subjects not previously trained on the objects ($n = 32$ total subjects; $n = 16$ for each object set) view a pair of object images, side by side, along with a scale numbered from 1 (very dissimilar) to 9 (very similar). Subjects performed two randomized block of 66 trials in which they based

their ratings on either *color* or *shape* similarity. We randomized trial order within both blocks, and counterbalanced block order of color or shape. Post-experiment debriefing revealed that the influence of the irrelevant feature on the current feature was minimal.

fMRI Procedure

Shape Retrieval Task: This task was only given to the 32 subjects who returned for the fifth and final session of the study. On each trial of the task, while undergoing fMRI, subjects read a question about one of the learned objects (e.g., "If you flipped a YERTS over, would it stand up straight?"). There were 20 questions (see Table 2), and each set of 20 questions was asked about each of the 12 objects, resulting in 240 total questions asked of the objects.

The trial structure was as follows: At the beginning of each trial, a "READY?" prompt appeared for 500 ms. A fixation cross then appeared for 500 ms, followed by the question about the object. While the question remained on the screen for 4500 ms, the subject was instructed to determine if the question referred to a plausible detail about the object's shape, responding "yes" or "no" via button press. At the end of the trial, a central fixation cross appeared for 500 ms, for a total trial duration of 6000 ms. Text was presented as white font on a black background.

Each subject completed four scanning runs of the shape retrieval task (approximately 9 minutes each) with 60 trials of the task per run. Using a rapid, event-related design, we presented a unique trial order to each subject, using Optseq2 (http://surfer.nmr.mgh.harvard.edu/optseq) to generate optimized pseudo-random stimulus presentation sequences. Experimental trials were intermixed with jittered fixation periods averaging six seconds in length.

After completing the shape retrieval task, all subjects completed the color perception functional localizer, which consisted of wheel-like visual stimuli that were made up of five smaller wedges, and that were either colored or grayscale (Figure 4, top left). On any given trial, participants fixated on a dash at the center of the wheel and indicated whether the wedges making up the wheel proceeded in order from lightest to darkest (i.e., a luminance judgment). The methods used for this localizer were identical to those used previously (Beauchamp et al., 1999; Hsu et al., 2011, 2012).

Image Acquisition—We acquired imaging data using a 3T Siemens Trio system with an 8-channel head coil and foam padding to secure the head in position. After we acquired T1-weighted anatomical images (TR = 1620 msec, TE = 3 msec, TI = 950 msec, voxel size = 0.9766 mm × 0.9766 mm × 1.000 mm), each subject performed the shape retrieval task, followed by the color perception task, while undergoing blood oxygen dependent (BOLD) imaging (Ogawa et al., 1993). We collected 870 sets of 42 slices using interleaved, gradient echo, echoplanar imaging (TR = 3000 msec, TE = 30 msec, FOV = 19.2 cm × 19.2 cm, voxel size = 3.0 mm × 3.0 mm × 3.0 mm). At least 9 seconds of "dummy" gradient and radio frequency pulses preceded each functional scan to allow for steady-state magnetization; no stimuli were presented and no fMRI data were collected during this initial time period.

Neuroimaging Data Analysis—We analyzed the data off-line using VoxBo (www.voxbo.org). Within VoxBo, we utilized a single scripting framework, which also called functions from SPM2 (http://www.fil.ion.ucl.ac.uk), and the FMRIB Software Library (FSL) toolkit (http://www.fmrib.ox.ac.uk/fsl). Anatomical data for each subject were processed using FSL to perform brain extraction (Smith, 2002), correct for spatial inhomogeneities (Zhang, Brady, & Smith, 2001) and to perform non-linear noise reduction (Smith & Brady, 1997). Using VoxBo, functional data were sinc interpolated in time to correct for the slice acquisition sequence, motion corrected with a six-parameter, least squares, rigid body realignment routine using the first functional image as a reference. We then used SPM2 to normalize to a standard template in Montreal Neurological Institute (MNI) space. Using VoxBo, the fMRI data were smoothed using a 9mm full-width half-max (FWHM) Gaussian smoothing kernel for univariate analyses, and with a 4mm smoothing kernel for multivariate analyses. With VoxBo, following preprocessing for each subject, a power spectrum for one functional run was fit with a 1/frequency function, and this model was used to estimate the intrinsic temporal autocorrelation of the functional data (Zarahn, Aguirre, & D'Esposito, 1997).

We fit a modified general linear model (Worsley & Friston, 1995) to each subject's data, in which task trials were each modeled as separate event with a 6 sec duration, and convolved with a standard hemodynamic response function. We included run effects (i.e., inter-run scanner drift) and movement spikes (i.e., TRs wherein we detected > 3.5 SD movement on a subject-by-subject basis) as covariates of no interest in the model. From this model, we computed parameter estimates for the task (compared to fixation baseline) at each voxel. These parameter estimates were included in the group-level random effects analyses described above.

Experiment 2

Participants—Twenty-four (n = 24) healthy subjects participated in the study (8 males, 16 females; average age = 24.1 years, range = 20 – 34 years). Twelve subjects participated in both experiments.

Materials and Procedure

Stimuli: We selected twelve object categories from two taxonomies: fruits/vegetables (FV) and household items (HHI). FV categories were: *apple, avocado, banana, beet, broccoli, carrot, cherry, lemon, lime, pumpkin, strawberry,* and *tomato*. HHI categories were: *clock, comb, fork, knife, ladle, nail file, scissors, spatula, spoon, tongs, toothbrush*, and *tweezers*. Just as the CS novel objects carried more diagnostic, associative color information relative to S objects, FV items carried more associative color information relative to HHI items. Note that FV and HHI taxonomies also differ in terms of relative diagnosticity (i.e., color versus other semantic features), since other non-color features also contribute to the overall representation.

In a parallel between-subjects design as described in Experiment 1, subjects were assigned to one of the two item sets. For the assigned item set, subjects performed the adjective generation and pairwise similarity tasks, though we only describe the results of the adjective

generation task here. When coding the descriptors from the adjective generation set, we included as color descriptors those that described surface properties (e.g., “shiny”). Unlike Experiment 1, for this experiment, we did not include the explicit color naming task, nor did an independent set of subjects perform the perceptual similarity task.

While subjects underwent fMRI, we used the same shape retrieval task from Experiment 1. Of the 20 original questions, we modified those questions that would be implausible for any of the familiar object categories, replacing them with appropriate, plausible questions (see Table 2). The fMRI task procedure, followed by functional localizers, was otherwise identical to that described in Experiment 1.

Experiment 1: Novel Objects - Results

Effects of feature diagnosticity on behavioral measures

Training Naming task—For naming task performance, we performed a mixed measures ANOVA (“color + shape” (CS) – $n = 29$; “shape” (S) – $n = 34$) on naming response accuracy, revealing a significant main effect of stimulus set ($F(1,61) = 11.42$, $p < 0.001$), a significant main effect of session ($F(1.28, 78.00) = 98.53$, $p < 0.001$), and a significant interaction of stimulus set and session ($F(1.28, 78.00) = 7.21$, $p < 0.01$). Critically, by the end of training, both groups were equally proficient at correctly producing the names of the learned objects (CS: M = 97.4% SE = 4.8%, S: M = 98.3%, SE = 2.9%; $t(61) = 1.25$, $p > 0.2$), such that any differences on subsequent tasks cannot be attributed to differences in how well both groups learned and knew the objects.

A similar ANOVA on RT revealed a significant main effect of stimulus set ($F(1,61) = 22.53$, $p < 0.001$), a significant main effect of session ($F(1.11, 67.80) = 72.49$, $p = 0.001$), but no interaction of stimulus set and session ($F(1.11, 67.80) = 1.49$, $p > 0.2$). By the fourth session, the groups significantly differed in RT, with CS subjects taking longer to produce the object names (average median RT for CS: 1417 ms, average median RT for S: 1077 ms; $t(61) = 5.39$, $p < 0.001$). The naming task results are shown in Figure 2.

The impact of feature diagnosticity on conceptual knowledge

We examined the effects of feature diagnosticity on three assessments of conceptual knowledge: (1) Did both groups of subjects learn the colors of the novel objects; (2) Did both groups of subjects prioritize color information equally; and (3) Did both groups of subjects use color information when considering the similarity of different objects to each other? When we asked subjects to identify the color of a named object, both groups could do so equally well (CS: $M = 93.4\%$, $SE = 2.3\%$, S: $M = 90.5\%$, $SE = 1.9\%$; $t(61) = 0.98$, $p > 0.3$, ns). However, when we asked subjects to describe the objects, CS subjects offered a (correct) color adjective as their first response nearly twice as often as did the S subjects (CS: $M = 87.9\%$, $SE = 4.0\%$, S: $M = 44.6\%$, $SE = 6.5\%$; $t(61) = 5.44$, $p < 0.001$) These results, shown in Figure 3, suggest that although the groups remembered object color equally, they did not prioritize color information equally.

We also found that CS and S subjects differed in how they used color information to evaluate the similarities of different objects to one another. Despite no explicit instruction to

base the similarity rating on any particular feature, critically, CS subjects assigned (on a 9-point scale) higher *general* similarity ratings to same-colored object pairs than did S subjects ($t(61) = 2.27$, $p = 0.03$). We observed a similar pattern when only comparing stimuli shared across both training groups. In the shared stimuli analysis, although the groups did not significantly differ in rating the same-colored pair (C+S: 4.9; S: 4.4; *t*-test across subjects: $t[31] = 0.99$; $p = 0.33$), they did judge the items in the five differently-colored pairs to be more dissimilar from each other (C+S: 1.5; S: 2.9; *t*-test across items: $t[8] = 7.03$, $p < .01$).

The impact of feature diagnosticity on neural representations

ROI univariate analysis—To establish functionally defined regions of interest (fROIs) in which we could assess any group differences in task effects, we first performed a group-level random effects analysis on the color perception data, comparing brain activity of colored stimuli to grayscale stimuli. This comparison is identical to previous work (Beauchamp, Haxby, Jennings, & DeYoe, 1999; Hsu et al., 2012, 2011; Simmons et al., 2007). No regions responded more to grayscale than colored stimuli. From the set of fROIs that emerged, we identified the peak cluster of voxels from posterior and anterior visual regions (identified as cuneus and fusiform gyrus). Both sets of regions (i.e., posterior and anterior) have been documented previously for their involvement in color perception and color knowledge retrieval (Beauchamp et al., 1999; Hsu et al., 2011; Martin, 2007; Simmons et al., 2007). To create fROIs of comparable size across regions, we identified approximately 50 maximally-responsive voxels in each region. Finally, within each of these fROIs, we calculated parameter estimates for each subject on the spatially-averaged time series across the 50 voxels in the fROI, using these parameter estimates to assess shape retrieval task effects (relative to fixation baseline) between groups. Critically, there were no group differences in RT for the shape retrieval task (C+S: average median RT: 2047 ms; S: average median RT: 2108 ms; $t(30) = 0.37$, $p > 0.7$). We used an independent samples *t*-test to assess the difference between groups.

Activation in the left fusiform region (48 voxels, peak voxel $t = 6.37$, Talairach coordinates: −30, −56, −17, BA 37) during the shape retrieval task was significantly greater for the CS subjects (mean % signal change = 0.41%, $SE = 0.07\%$) than for the S subjects (mean % signal change = 0.22%, $SE = 0.06\%$; $t(30) = 2.02$, $p = 0.05$, see Figure 4). Performing the same analysis with individually-defined ROIs yielded a similar pattern. In contrast, activation in the cuneus region (52 voxels, peak voxel $t = 9.56$, Talairach coordinates: 3, −92, 20, BA 19/17) did not show a significant group difference in activity during the shape task (CS: mean % signal change = 0.32%, $SE = 0.08\%$; S: mean % signal change = 0.19%, $SE = 0.08\%$; $t(30) = 1.11$, $p > 0.2$, ns). The region by group interaction was not significant.

In order to rule out a task difficulty explanation (i.e., attributing greater fusiform activity to the task being harder for CS participants), we examined “accuracy” on the memory task. Although there were no “correct” answers for the shape questions, we derived a consensus measure for each question by counting the number of “yes” and “no” responses, calculating the absolute value of their difference, and dividing by the total number of responses. Lower consensus values would approach 0, whereas higher consensus would approach 1. If CS subjects found the task more difficult, a task difficulty hypothesis would predict lower

consensus on their answers. However, CS subjects had *higher* consensus than S subjects on the memory task (CS: $M = 0.66$, $SE = 0.02$; S: $M = 0.55$, $SE = 0.02$, $t(478) = 4.05$, $p < 0.001$).

Multivariate neural similarity analysis—We next adopted a measure of neural similarity from Weber and colleagues (2009) in order to see if activation patterns in the left fusiform fROI (48 voxels) predicted behavioral similarity ratings. First, we pre-processed data for this analysis with a smaller smoothing kernel (4 mm, rather than 9 mm) than for the univariate analyses, as larger smoothing kernels can be destructive for multivariate analyses. Then, for each item, we identified a pattern of activation of vector length equal to the number of voxels in the fROI. Although voxel order in the vector was arbitrary, it remained consistent across all patterns. Some voxels within the fROI were, on average, more active than others; thus, in order to prevent mean activation of voxels from driving our similarity measure (a Pearson correlation of neural similarity), we mean-centered each voxel's response to its average response across all items. We calculated neural similarity by correlating each of the 66 vector pairs (averaged over subjects), and then assessed whether these values could predict two sets of behavioral ratings of similarity: the general similarity ratings obtained by the subjects (by memory), as well as the similarity ratings obtained by an independent group of subjects (by perception) (see Methods: Untrained Similarity Rating for details on obtaining feature-specific behavioral similarity ratings).

We conducted these analyses with both sets of behavioral similarity ratings for specific reasons. First, we used the logic behind theories of embodied cognition - namely, that color-sensitive brain systems are recruited when thinking about color - as the motivation for our decision to use perceptual similarity judgments from an independent group of participants. Specifically, we wanted perceptual judgments from an untrained set of participants for two reasons – a) previous knowledge about the objects (e.g., object name) would not influence the similarity ratings, and b) we could probe participants on specific and critical perceptual features (i.e., color or shape). We could then correlate this relatively *clean* set of perceptual similarity ratings with the neural similarity data. Second, since we were also interested in correlating the neural similarity data with the (memory-based) general similarity data from the trained participants (i.e., a test for within-subjects similarity correlations), we used this second set of behavioral ratings as well.

Because we could not assume a linear relationship for the behavioral ratings of similarity, we used the Spearman rank correlation coefficient to assess the relationship between neural and behavioral similarity. Finally, we ran a Monte Carlo simulation in order to arrive at the appropriate *p*-values for the similarity correlations.

Perceptual Similarity: As shown in Figure 5, color similarity ratings approached significance in predicting neural similarity for the CS subjects ($r_s = 0.23$, $p = 0.06$; 95% CI: −0.01 to 0.45), but not for the S subjects ($r_s = -0.17$, $p = -0.18$; 95% CI: −0.39 to 0.08). These predictions were significantly different from each other ($Z = 2.27$, $p < 0.05$). Shape similarity ratings did not predict neural similarity in this region for either group (CS: $r_s = 0.19$ $p = 0.13$; S: $r_s = -0.02$, $p > 0.8$;), and the two groups did not differ from each other ($Z = 1.17$; $p > 0.2$).

General Similarity: In the localizer-defined left fusiform gyrus ROI that showed a group effect, we correlated the average behavioral *general* similarity rating with the average neural similarity measure. The general similarity ratings predicted neural similarity for CS subjects (r_s = 0.29, p = 0.02; 95% CI: 0.05 to 0.50) and not for S subjects (r_s = −.07, p = 0.57; 95% CI: −0.31 to 0.18), and the correlations were significantly different from each other (Z = 2.06, p = 0.04).

Exploratory analyses—To assess the specificity of our effect, we examined other regions of the left ventral temporal cortex other than those used for our primary *a priori* analyses. In line with previous work, we expanded our search to left ventral temporal cortex in line with left-lateralized brain regions involved in knowledge retrieval (cf. Martin et al., 1995; Chao et al., 1999). In an anatomically-defined left ventral temporal cortex region (~ 5500 voxels), we looked for voxel clusters (> 50 voxels) showing a task (task versus baseline) X group (color+shape versus shape) interaction at a cluster-corrected, permuted threshold of $\alpha < 0.05$ (t = 2.92). Only the left inferior temporal gyrus (Talairach coordinates of peak voxel: −56, −53, −12, BA 20), surpassed this threshold, both within the anatomically defined region, and when we unmasked the rest of the brain to examine whether other regions demonstrated this interaction.

Here (see the bottom panel of Figure 4), we found significantly greater activity during the shape retrieval task for CS subjects than for S subjects (note: the means plotted in the lower panel of this figure are intended to provide descriptive data about the ROI, not an independent inferential test, as they are taken from the voxels identified as having a reliable interaction in the Exploratory analysis). Moreover, as shown in Figure 6, we also found that the extent to which subjects prioritized color during the adjective generation task (i.e., how often they listed object color first) predicted activity in this region (C+S: r = −0.18, p = 0.49; S: r = 0.30, p = 0.24, combined: r = 0.50, p < 0.01). We observed similar trends in the left fusiform gyrus (C+S: r = 0.21, p = 0.42; S: r = −0.07, p = 0.79; combined: r = 0.30, p = 0.09) and the cuneus (C+S: r = 0.19, p = 0.47; S: r = 0.09, p = 0.73; combined: r = 0.23, p = 0.20), the two regions identified from the color perception functional localizer. This result suggests that during object knowledge retrieval, diagnostic features may be automatically activated.

Finally, in a second exploratory analysis, we assessed the specificity of the shape retrieval task effect (relative to baseline) by conducting a whole-brain analysis. Using a permuted threshold (t > 6.22; $\alpha < 0.005$), we identified the local maxima that surpassed this threshold and derived the corresponding brain regions, which are now reported in Table 3. As seen in the table, activation focuses on color-selective regions (e.g., left fusiform gyrus, left lingual gyrus, cuneus, precuneus) in addition to other regions.

Experiment 2: Familiar Objects - Results

Effects of familiar object feature diagnosticity on behavioral measures

Adjective Generation Task—A repeated measures ANOVA revealed significant main effects of both condition (F(1,21) = 21.14, p < 0.001) and adjective order (F(1,21) = 13.56, p < 0.001), but no interaction (F(1,21) = 0.32, p > 0.5). Even when including surface

descriptors in the adjectives as colors, such as “shiny,” “metallic,” “plastic,” and “wooden,” subjects describing FV items listed color first more often than subjects describing HHI items (FV: $M = 54.9\%$, $SE = 10.3\%$; HHI: $M = 24.2\%$, $SE = 5.4\%$; $t(21) = 2.56$, $p < 0.02$). These results are comparable to Experiment 1, supporting the idea that within the context of this particular task, both novel and familiar object categories share commonalities.

The generalization of feature diagnosticity effects to familiar object categories

One of the main strengths of novel object studies (i.e., experimenter-manipulated control) is also a constraint: it is often unclear to what extent the results will generalize to familiar objects for which there is natural variation in stimulus characteristics. Thus, Experiment 2 asked whether familiar, real-world objects that varied in relatively diagnostic color association (i.e., fruits and vegetables versus household items) would yield findings in line with those of Experiment 1. In an item-based analysis using all three ROIs (i.e., the functional color localizer-identified fusiform gyrus and cuneus; the exploratory analysis-identified inferior temporal gyrus) from Experiment 1, we compared item responses from both experiments as a function of color prioritization. For all 48 items, within each ROI from Experiment 1, we obtained the response to each individual item across all 20 questions, averaged across all subjects. This analysis allows us to compare item responses across conditions and across experiments.

As observed in Figure 7, color prioritization varied across both experiments. Across conditions, the distribution of items does not overlap for novel objects, but does for familiar objects. In the cuneus, comparing item responses across the two common object categories reveals significantly greater percent signal change for FV items ($M = 0.40$, $SE = 0.03$) relative to HHI items ($M = 0.28$, $SE = 0.01$; $t(22) = 3.79$, $p < 0.007$). Because the FV items are color-associated, but the HHI items tend to not be color-associated, this result parallels previously reported chromaticity effects in memory (Hsu et al., 2012). Notably, the same pattern was also observed in the left fusiform gyrus, the region where we had discovered a group effect of feature diagnosticity in Experiment 1 (FV: mean % signal change = 0.26; $SE = 0.01$; HHI: mean % signal change = 0.22; $SE = 0.01$; $t(22) = 2.25$, $p < 0.05$). Further, both regions demonstrated positive correlations between color prioritization and BOLD signal in Experiment 2 (cuneus: $r = 0.58$, $p = .003$; fusiform: $r = 0.44$, $p = 0.03$). The latter fusiform region replicates a similar pattern observed in Experiment 1 (fusiform: $r = 0.83$, $p < 0.001$), but not in the cuneus ($r = 0.28$, $p = 0.18$). That color prioritization predicted responses for both novel and common object categories in the left fusiform gyrus suggests some commonalities in the relative role of feature diagnosticity regardless of stimuli type. Interestingly, we observed different patterns in the left inferior temporal gyrus across experiments. Whereas color prioritization positively correlated with BOLD signal for items in Experiment 1 ($r = 0.87$, $p < 0.001$), color prioritization negatively correlated with BOLD signal for items in Experiment 2 ($r = -0.46$, $p < 0.05$). We discuss possible reasons for this divergence in the General Discussion.

General Discussion

We report several behavioral and neural results indicating that feature diagnosticity affects concept representations. In ventral temporal cortex, and specifically in the left fusiform

gyrus and left inferior temporal gyrus, we found greater activity for subjects who had learned color was a useful, diagnostic feature when performing a task that did *not* explicitly require color retrieval. We also found that behavioral ratings of color similarity predicted neural similarity for CS subjects only, and that color prioritization predicted activity in color-selective fusiform gyrus; this latter effect was also evident in the set of familiar objects. Together, these results provide evidence that the behavioral effects of feature diagnosticity (measured at least in part by color prioritization) arise from varying degrees of automatic recruitment of the diagnostic feature; this brain-behavior correlation was evident across both stimulus sets. To our knowledge, this study is the first to explain rather than describe the importance of feature diagnosticity, both when the diagnostic feature (here, color) is systematically manipulated, and in a more familiar real-world context.

Although both training groups were equally able to identify the color of the object when explicitly asked to, the CS subjects listed color first more frequently when naming features. This result is particularly interesting in light of some previous work (Connolly, Gleitman, & Thompson-Schill, 2007), which used an implicit similarity measure to demonstrate that although both sighted and congenitally blind subjects were equally proficient at *knowing* the colors of fruits and vegetables, only sighted subjects *used* color as the primary basis for their judgment. The authors suggested that visual experience (or lack thereof) had contributed to a fundamental group difference in how conceptual representations for these categories were structured. Our design matched the training stimuli in terms of color uniqueness and probability of occurrence (i.e., for both sets, it was always the case that $P(\text{object} \mid \text{color}) = 0.5$). Despite a fixed level of absolute diagnosticity, we found fundamental differences in how subjects used color, which was relatively more diagnostic (compared to shape). We can stipulate color knowledge of a *klarve* for both groups of subjects from the color naming task, but the adjective generation task yields information about the *usefulness* of color in distinguishing a *klarve* from the other objects in the set.

Further, the shapes were deliberately created such that they bore no resemblance to familiar objects and thus would not be easily named. Subjects tried – and often struggled – to generate descriptors, sometimes resorting to shape adjectives that were easily verbalized (e.g., a *klarve* is “curved”), or likening shapes to ones that they knew (e.g., a *klarve* is “football-like”). Given this observed difficulty, one might have predicted the subjects to produce color descriptors, which *are* easily named. Despite this difference in likely ease of production, S subjects did not produce color descriptors before non-color descriptors, and, for some S subjects, color was never mentioned at all. This result strengthens our argument that the object set differences, together with subsequent differences in visual experience, contributed to fundamental differences in how the groups represented the novel objects.

The pairwise similarity data provided a complementary method for investigating conceptual knowledge; according to some theories of concepts, similarity amongst instances of a category is critical for category (Murphy, 2004). The data here demonstrate a fundamental difference in how the CS subjects considered the general similarity of same- versus different-color object pairs. Given the unavoidable heterogeneity in constructing the two object sets, restricting the analysis to shared stimuli between the groups (*klarve, hinch, fulch, screll)* replicated our initial findings (specifically in terms of dissimilarity),

demonstrating that diagnostic features can be regarded in the context of long-term experience with other objects in the set. Not only does use of feature knowledge affect a conceptual representation, but our data show that the learned context of the objects can also affect conceptual representations.

Turning to the neuroimaging data, we hypothesized a group difference in accessing color as a diagnostic feature during a shape retrieval task in the left fusiform gyrus. Our findings were in line with our initial hypothesis, in that the left fusiform gyrus, which is known to be a region involved in color perception (Beauchamp et al., 1999; Hsu et al., 2011; Simmons et al., 2007), was indeed more active during the shape retrieval task for CS subjects than for S subjects. These results are all the more compelling given than the shape retrieval task never explicitly probed subjects about object color. In fact, color was irrelevant to the task. This result, suggesting automatic retrieval of diagnostic features even when retrieving other object features, is consistent with temporal information revealed in a related event-related potential (ERP) study: subjects categorizing novel objects showed ERP patterns as early as 117 ms when remembering diagnostic features of the learned objects. However, this early effect was only seen in occipitoparietal electrodes when subjects had pantomimed actions with the novel objects, rather than pointing to them (Kiefer et al., 2007). Though we did not find a significant group effect in the cuneus region (i.e., the other region identified in the functional color localizer), the results were numerically in the same direction. This was a slightly surprising but not an undocumented finding, as previous work has demonstrated differential activation of color perception in posterior versus anterior regions (e.g., Beauchamp et al., 1999). Finally, our multivariate analyses revealed that neural similarity of patterns in left fusiform gyrus were linked to general (i.e., from trained subjects) and perceptual (i.e., from untrained subjects) similarities for the CS subjects only (that is, when color had relatively high diagnosticity). Since color did not yield the same relatively diagnostic information for S objects, this may explain the elimination of the correlation between behavioral and similarity for these subjects.

Further, our follow-up exploratory analyses revealed an unexpected pattern in the left inferior temporal gyrus. Previous work has shown this region - lateral and anterior to the medial fusiform region - to be involved in color knowledge retrieval; it is more active when subjects name colors (of achromatically presented object drawings) than when they name the objects themselves (Chao & Martin, 1999). We find that color prioritization is correlated with activity in this region (with a similar pattern in other color perception regions). This brain-behavior correlation indicates that the behavioral effect of feature diagnosticity arises from differing degrees of automatic recruitment of color information. However, we wish to mention one caveat to this particular exploratory analysis. Specifically, the within-group assessments of this data were not significant, and effects only emerged when combining the groups together; yet, within-group assessments are challenging. That is, it was difficult for us to assess the same relationship within both groups because the two groups demonstrated differing patterns of data distribution in terms of color prioritization. Within the S group, the correlation was positive ($r = 0.30$) – while not significant, it was numerically in the direction that we would expect (given values that are a bit more normally distributed), but might be under-powered. Despite this caveat, we note that the present study is unusual in the fMRI literature in that it involves a between-subjects manipulation, which has considerably less

power than more typical within-subjects designs. The between-subjects design was a necessity in this case, but one consequence is that we had less statistical power than desired. We urge readers to consider the overall body of conclusions reported here as a broader demonstration of how behavioral effects of feature diagnosticity can arise from the recruitment of color information.

In considering the results across all three regions, the composition of the novel object sets, in a sense, forced subjects to categorize objects according to strict color-shape conjunctions. Thus, our experimental design may have been more amenable towards group differences in an anterior region involved in feature integration, but it does not preclude similar group differences in a posterior color region. Group differences in the shape task showed the same directional effect in the cuneus as in fusiform gyrus, though not significantly so. Since the magnitude of this difference increased from posterior to anterior regions of ventral temporal cortex, this result suggests an increased sensitivity to diagnostic features in regions tuned to object categorization. In line with this theory, macaque IT cortex differentially responded to diagnostic features along a posterior-anterior axis, with only the anterior portion of the recording area responded with diagnostic local field potential (LFP) activity (Nielsen et al., 2006).

Finally, Experiment 2 allowed us to compare between novel and familiar object categories. In both color perception regions, color prioritization correlated with brain activity during the shape task, suggesting automatic retrieval of diagnostic features, and in particular, a neural basis for the taxonomy differences in relative color association. The reversal of this trend in the left inferior temporal gyrus in Experiment 2 suggests a different role for this region in concept representations. One possibility is that the region represents the contribution of an object feature, in light of all other known object features. Whereas color information constituted 50% of the features known about novel object categories, it likely constituted a smaller percentage of all known features of HHI object categories (where other features might include function, texture, etc.), and an even smaller percentage for FV object categories (where other feature might include taste, size, etc). However, there may be other explanations that better explain this seemingly contradictory reversal in correlation patterns across both experiments, and future research should address this finding. Taken together, this comparison demonstrates the utility of training studies, in that they allow amplification of an otherwise muddied gradient of information that is of interest.

We have argued here that the retrieval of diagnostic features can automatically activate color-selective brain regions, but some of these brain regions - particularly the fusiform gyrus - are also involved in shape processing (e.g., Bar et al., 2001; Gerlach, Law, & Paulson, 2006; Op de Beeck, Béatse, Wagemans, Sunaert, & Van Hecke, 2000). As such, there remains the possibility that the diagnostic feature in question may not have been color per se, but the conjunction of shape and color. The data in our study cannot rule out this possibility. However, a recent MVPA study (Coutanche & Thompson-Schill, in press) demonstrated that within a region often associated with color processing, a classifier could decode meaningful information about color and about shape, but could not decode the conjunction of the two features (in that study, only in the anterior temporal lobe could the classifier decode feature conjunction information). We believe that this result makes it less

likely for a color processing region to carry feature conjunction information, but this is an intriguing idea that the field should pursue further. That is, which brain regions carry information about independent conceptual features, and which brain regions carry information about conjoint features? One promising method for addressing this question is to compare metrics that measure independent (i.e., city-block) or conjoint (i.e., Euclidean) featural information (Drucker, Kerr, & Aguirre, 2009). Future explorations of conjunctive feature coding might also benefit from potential links to the literature on learning of conjunctive versus non-conjunctive rules in categorization tasks (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Maddox, 2011; Ell, Weinstein, & Ivry, 2010). Finally, we emphasize that whether a singular feature or conjunction of multiple features is the diagnostic component of the concept, in either case, we observe automatic retrieval of information that is seemingly task-irrelevant, but only in cases when that information is diagnostic.

Collectively, the results of the current work are the first to provide a neural explanation for the behavioral effects of feature diagnosticity, namely that these effects arise from automatic recruitment of the diagnostic information. Our findings suggest that neural representations may not be stable and fixed, but may instead be far more flexible than previously thought (Binder & Desai, 2011; Hoenig, Sim, Bochev, Herrnberger, & Kiefer, 2008; Kiefer & Pulvermüller, 2011). More broadly, our results point to the notion that feature diagnosticity is one of many sources contributing to variation in concept representations, the neural bases of which underlie our ability to describe *and* define characteristics of the massive variety of objects that we encounter on a daily basis.

Acknowledgments

This work was funded by R01-MH070850 to S.L.T.-S. and F31-AG034743 to N.S.H. We thank Matt Weber for help with data analysis, Emily Kalenik and Lauren Hendrix for help with data collection, members of the Thompson-Schill lab for generous feedback and discussion, and two anonymous reviewers for comments on an earlier version of this manuscript.

References

Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM. A Neuropsychological Theory of Multiple Systems in Category Learning. Psychological Review. 1998; 105(3):442–481. [PubMed: 9697427]

Ashby FG, Maddox WT. Human category learning 2.0: Human category learning 2.0. Annals of the New York Academy of Sciences. 2011; 1224(1):147–161. [PubMed: 21182535]

Bar M, Tootell RB, Schacter DL, Greve DN, Fischl B, Mendola JD, Dale AM. Cortical mechanisms specific to explicit visual object recognition. Neuron. 2001; 29(2):529–535. [PubMed: 11239441]

Beauchamp MS, Haxby JV, Jennings JE, DeYoe EA. An fMRI version of the Farnsworth-Munsell 100-Hue test reveals multiple color-selective areas in human ventral occipitotemporal cortex. Cerebral Cortex (New York, NY : 1991). 1999; 9(3):257–263.

Binder JR, Desai RH. The neurobiology of semantic memory. Trends in Cognitive Sciences. 2011; 15(11):527–536. [PubMed: 22001867]

Chao LL, Martin A. Cortical regions associated with perceiving, naming, and knowing about colors. Journal of Cognitive Neuroscience. 1999; 11(1):25–35. [PubMed: 9950712]

Connolly AC, Gleitman LR, Thompson-Schill SL. Effect of congenital blindness on the semantic representation of some everyday concepts. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104(20):8241–8246. [PubMed: 17483447]

Coutanche M, Thompson-Schill S. Creating concepts from converging features in human cortex. Cerebral Cortex. (in press).

Cox DD, Savoy RL. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage. 2003; 19(2): 261–270. [PubMed: 12814577]

Cree GS, McNorgan C, McRae K. Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2006; 32(4):643–658.

Drucker DM, Kerr WT, Aguirre GK. Distinguishing Conjoint and Independent Neural Tuning for Stimulus Features With fMRI Adaptation. Journal of Neurophysiology. 2009; 101(6):3310–3324. [PubMed: 19357342]

Ell SW, Weinstein A, Ivry RB. Rule-based categorization deficits in focal basal ganglia lesion and Parkinson's disease patients. Neuropsychologia. 2010; 48(10):2974–2986. [PubMed: 20600196]

Gerlach C, Law I, Paulson OB. Shape configuration and category-specificity. Neuropsychologia. 2006; 44(7):1247–1260. [PubMed: 16289641]

Grossman E, Blake R, Kim C. Learning to see biological motion: Brain activity parallels behavior. Journal of Cognitive Neuroscience. 2004; 16:1669–1679. [PubMed: 15601527]

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. Science. 2001; 293(5539): 2425–2430. [PubMed: 11577229]

Hoenig K, Sim EJ, Bochev V, Herrnberger B, Kiefer M. Conceptual flexibility in the human brain: dynamic recruitment of semantic maps from visual, motor, and motion-related areas. Journal of Cognitive Neuroscience. 2008; 20(10):1799–1814. [PubMed: 18370598]

Hsu NS, Frankland SM, Thompson-Schill SL. Chromaticity of color perception and object color knowledge. Neuropsychologia. 2012; 50(2):327–333. [PubMed: 22192637]

Hsu NS, Kraemer DJM, Oliver RT, Schlichting ML, Thompson-Schill SL. Color, context, and cognitive style: variations in color knowledge retrieval as a function of task and subject variables. Journal of Cognitive Neuroscience. 2011; 23(9):2544–2557. [PubMed: 21265605]

James TW, Gauthier I. Auditory and action semantic features activate sensory-specific perceptual brain regions. Current Biology: CB. 2003; 13(20):1792–1796. [PubMed: 14561404]

Jimura K, Poldrack RA. Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia. 2012; 50(4):544–552. [PubMed: 22100534]

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

Kiefer M, Pulvermüller F. Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. Cortex; a Journal Devoted to the Study of the Nervous System and Behavior. 2011

Kiefer M, Sim EJ, Liebich S, Hauk O, Tanaka J. Experience-dependent plasticity of conceptual representations in human sensory-motor areas. Journal of Cognitive Neuroscience. 2007; 19(3): 525–542. [PubMed: 17335399]

Martin A. The representation of object concepts in the brain. Annual Review of Psychology. 2007; 58:25–45.

Martin A, Haxby JV, Lalonde FM, Wiggs CL, Ungerleider LG. Discrete cortical regions associated with knowledge of color and knowledge of action. Science (New York, NY). 1995; 270(5233): 102–105.

Murphy, G. The big book of concepts. Cambridge, MA: MIT Press; 2004.

Nielsen KJ, Logothetis NK, Rainer G. Dissociation Between Local Field Potentials and Spiking Activity in Macaque Inferior Temporal Cortex Reveals Diagnosticity-Based Encoding of Complex Objects. The Journal of Neuroscience. 2006; 26(38):9639–9645. [PubMed: 16988034]

O'Toole AJ, Jiang F, Abdi H, Haxby JV. Partially Distributed Representations of Objects and Faces in Ventral Temporal Cortex. Journal of Cognitive Neuroscience. 2005; 17(4):580–590. [PubMed: 15829079]

Ogawa S, Menon RS, Tank DW, Kim SG, Merkle H, Ellermann JM, Ugurbil K. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A

comparison of signal characteristics with a biophysical model. Biophysical Journal. 1993; 64(3): 803–812. [PubMed: 8386018]

Op de Beeck H, Béatse E, Wagemans J, Sunaert S, Van Hecke P. The Representation of Shape in the Context of Visual Object Categorization Tasks. NeuroImage. 2000; 12(1):28–40. [PubMed: 10875900]

Op de Beeck HP, Torfs K, Wagemans J. Perceived Shape Similarity among Unfamiliar Objects and the Organization of the Human Object Vision Pathway. The Journal of Neuroscience. 2008; 28(40):10111–10123. [PubMed: 18829969]

Polyn SM, Natu VS, Cohen JD, Norman KA. Category-Specific Cortical Activity Precedes Retrieval During Memory Search. Science. 2005; 310(5756):1963–1966. [PubMed: 16373577]

Rastle K, Harrington J, Coltheart M. 358,534 nonwords: The ARC Nonword Database. Quarterly Journal of Experimental Psychology. 2002; 55A:1339–1362. [PubMed: 12420998]

Schyns P. Diagnostic recognition: Task constraints, object information, and their interactions. Cognition. 1998; 67:147–179. [PubMed: 9735539]

Sigala N, Logothetis NK. Visual categorization shapes feature selectivity in the primate temporal cortex. Nature. 2002; 415(6869):318–320. [PubMed: 11797008]

Simmons W, Ramjee V, Beauchamp M, McRae K, Martin A, Barsalou L. A common neural substrate for perceiving and knowing about color. Neuropsychologia. 2007; 45(12):2802–10. [PubMed: 17575989]

Smith LB, Jones SS, Landau B, Gershkoff-Stowe L, Samuelson L. Object name learning provides on-the-job training for attention. Psychological Science. 2002; 13(1):13–19. [PubMed: 11892773]

Smith SM. Fast robust automated brain extraction. Human Brain Mapping. 2002; 17(3):143–155. [PubMed: 12391568]

Smith SM, Brady JM. SUSAN - a new approach to low level image processing. International Journal of Computer Vision. 1997; 23(1):45–78.

Tanaka JW, Presnell LM. Color diagnosticity in object recognition. Perception & Psychophysics. 1999; 61(6):1140–1153. [PubMed: 10497433]

Weber M, Thompson-Schill SL, Osherson D, Haxby J, Parsons L. Predicting judged similarity of natural categories from their neural representations. Neuropsychologia. 2009; 47(3):859–868. [PubMed: 19162048]

Weisberg J, van Turennout M, Martin A. A neural system for learning about object function. Cerebral Cortex. 2007; 17:513–521. [PubMed: 16581980]

Worsley KJ, Friston KJ. Analysis of fMRI time-series revisited--again. NeuroImage. 1995; 2(3):173–181.10.1006/nimg.1995.1023 [PubMed: 9343600]

Zarahn E, Aguirre GK, D'Esposito M. Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. NeuroImage. 1997; 5(3):179–197. [PubMed: 9345548]

Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imag. 2001; 20(1):45–57.

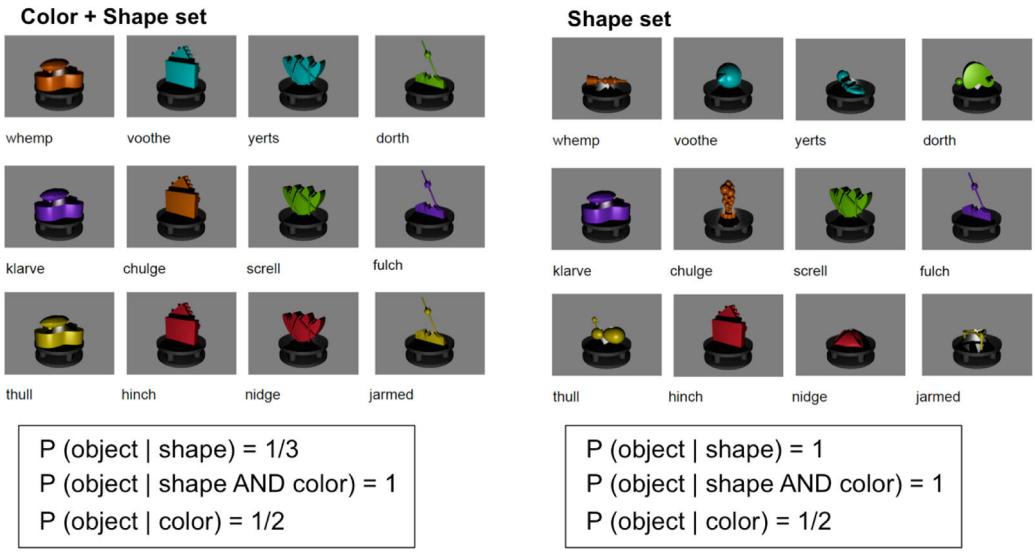

Figure 1.
Exemplars from the CS and S objects, demonstrating set differences.

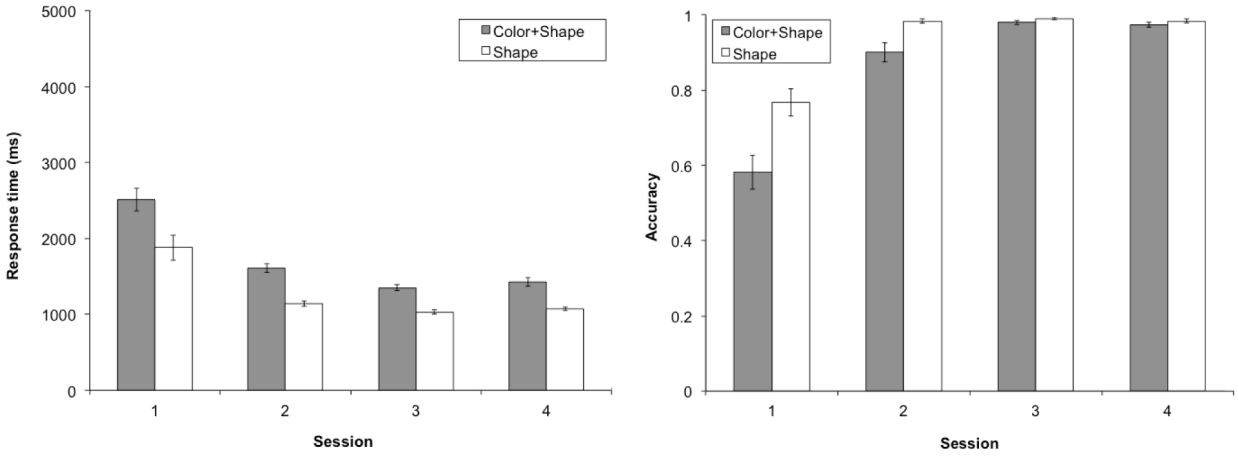

Figure 2. Behavioral performance on the naming task across training sessions

Response time (left) and accuracy (right) performance are shown for both groups. The groups did not differ in accuracy by the end of training.

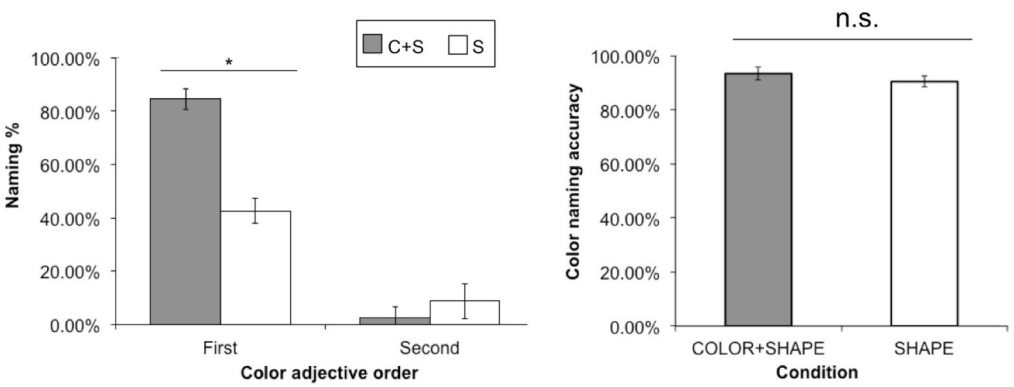

Figure 3. The training groups differ in prioritization of color information

(left) As measured by the frequency of listing color early in an adjective generation task, CS subjects listed color as the first adjective earlier and more often than did S subjects. (right) In contrast, the groups could identify colors of the novel objects equally well when explicitly asked.

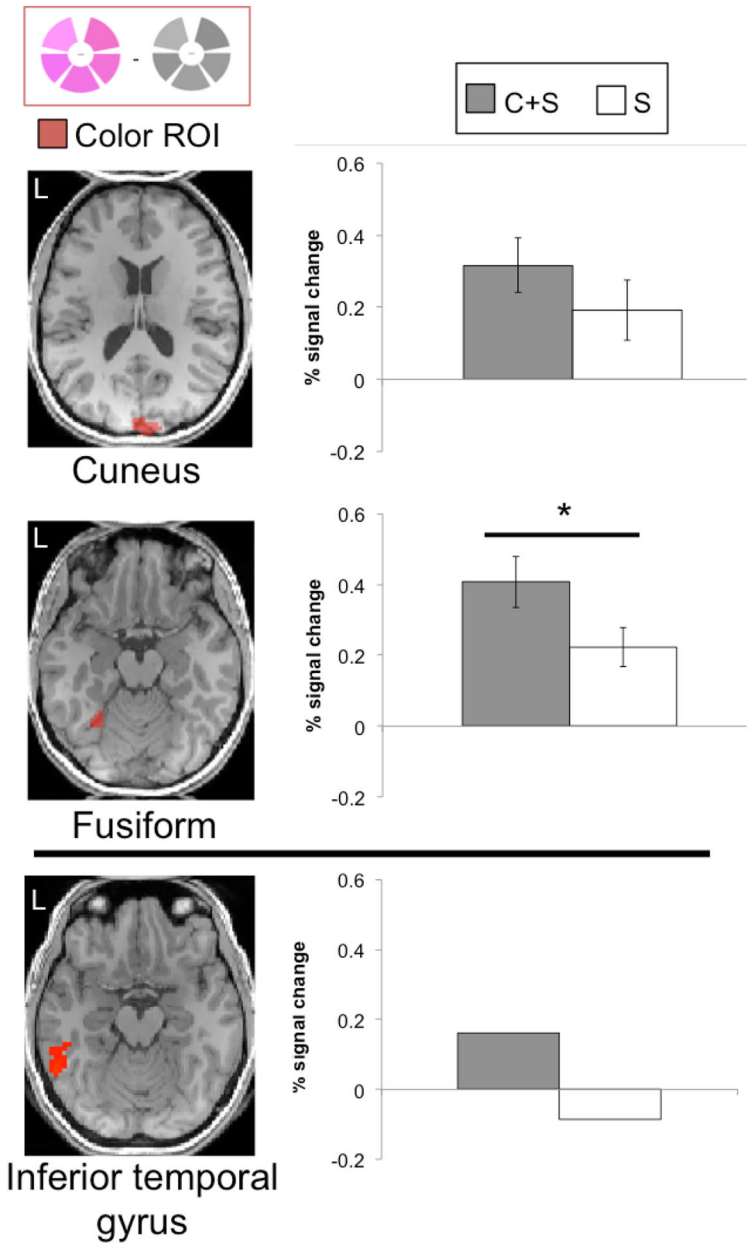

Figure 4. Retrieval of a diagnostic feature automatically activates color-sensitive regions in ventral temporal cortex

During a shape retrieval task, the left fusiform gyrus, a region involved in color perception as defined by greater response to chromatic than achromatic visual stimuli, was more active for CS subjects than for S subjects. The cuneus region showed a similar pattern that did not reach significance. Exploratory analyses revealed that the left inferior temporal gyrus demonstrated a significant task x group interaction, with CS subjects demonstrating more task activity than S subjects (note: the means plotted in the lower panel of this figure are intended to provide descriptive data about the ROI, not an independent inferential test, as

they are taken from the voxels identified as having a reliable interaction in the Exploratory Analysis).

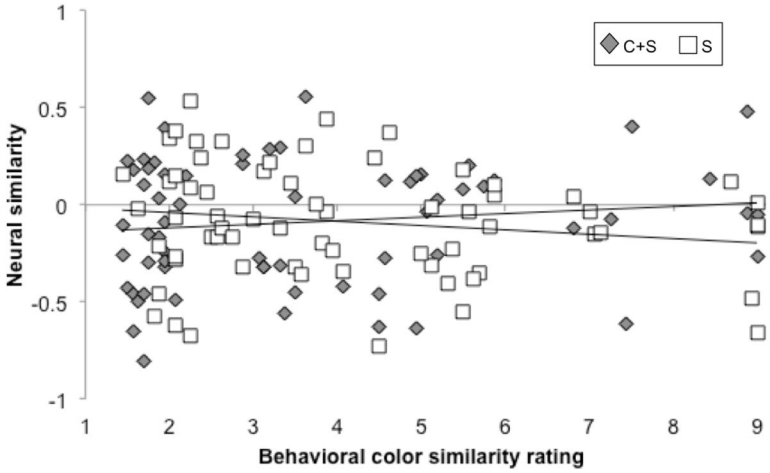

Figure 5. Behavioral color similarity predicts neural similarity in the left fusiform gyrus

Behavioral ratings of color similarity (derived from a set of untrained subjects) approach significance in predicting neural similarity of novel object activation patterns in the left fusiform gyrus, but only for the CS subjects, shown in gray. S subjects are shown in white. Each data point represents a pairwise combination of novel objects, averaged across all subjects.

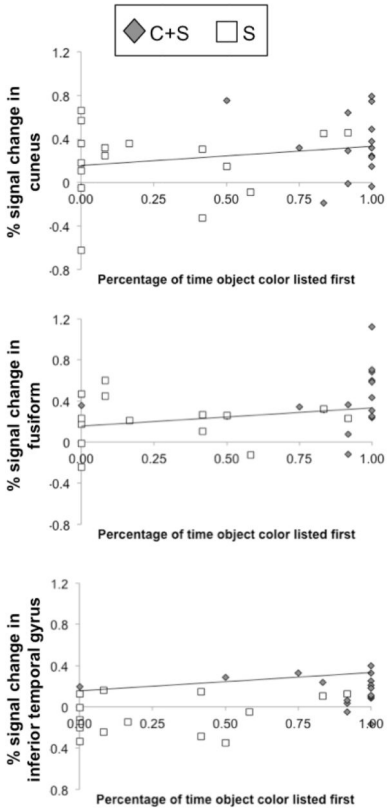

Figure 6. Color prioritization predicts task activity in ventral regions

Prioritizing color during the adjective generation task only correlated significantly with activity in the left inferior temporal gyrus, a second region active during the shape retrieval task that was identified through secondary exploratory analyses. Patterns in the same direction were observed in the left fusiform gyrus and cuneus. Each data point represents the BOLD response from a given subject, averaged across all items.

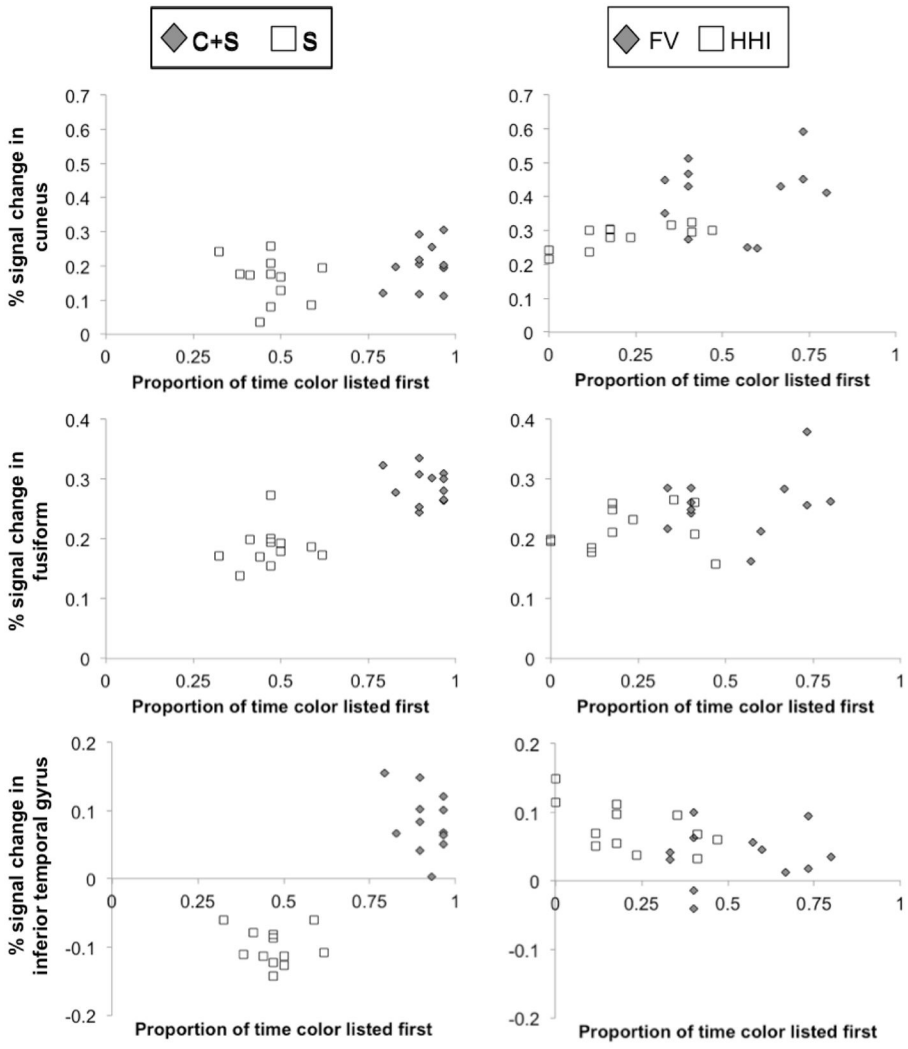

Figure 7. Item-based analyses reveal that feature diagnosticity effects generalize across stimulus sets

Across the three ROIs (posterior cuneus and anterior fusiform identified from the color perception localizer, inferior temporal gyrus identified from the exploratory analysis), we analyzed averaged item-level responses in signal change and color prioritization for Experiment 1 (left) and Experiment 2 (right). Each data point represents the BOLD response to a given item, averaged across all subjects.

Table 1

Subject training schedule, in which the specific combination of tasks is indicated for each session of the experiment.

	Session				
	1	**2**	**3**	**4**	**5**
Training					
Training Videos	◆	◆	◆		
Naming Task	◆	◆	◆	◆	
Testing					
Adjective Generation			◆		
Pairwise Similarity				◆	
Explicit Color Naming				◆	
fMRI + memory task					◆

Table 2

The questions used in the shape retrieval task that subjects performed while undergoing fMRI. “Chulge” and “spoon” here are example items – we asked the same list of 20 questions of all 12 object categories in each experiment. The lists differ slightly across experiments in order to maintain plausibility. As such, for Experiment 2, we replaced questions from Experiment 1 that we did not think were suitable for the objects in Experiment 2. Those questions are indicated by **.

List of Questions for Experiment 1:
Could you cut something with a CHULGE?
**Could you poke a hole with a CHULGE?
Could you roll a CHULGE down a hill?
Could you use a CHULGE as a weapon?
Does a CHULGE have corners?
**If you flipped a CHULGE over, would it stand up straight?
**Is a CHULGE bulging?
Is a CHULGE bumpy?
**Is a CHULGE cubic?
Is a CHULGE flimsy?
Is a CHULGE fragile?
Is a CHULGE made up of smaller parts?
Is a CHULGE rounded?
Is a CHULGE sharp?
Is a CHULGE symmetrical?
**Is a CHULGE tied together?
Would a CHULGE be easy to wrap up (e.g., as a present)?
Would you be able to spin a CHULGE?
Would you call a CHULGE curved?
**Would you consider a CHULGE to be flat?
List of Replacement Questions for Experiment 2:
**Could you poke a hole in a piece of paper with a SPOON?
**Does a SPOON have any protrusions from its main body?
**If you flipped a SPOON upside down, would it stand up straight?
**Is a SPOON smooth?
**Is a SPOON square?
**Would you be able to pinch a SPOON with two fingers?

Table 3

Regions identified from the whole-brain, permuted analysis of the shape retrieval task. Coordinates are in Talairach space and are given for the peak voxel (local maximum) with corresponding *t*-value. Note that these *t*-values correspond to regions identified in the shape retrieval task, whereas the *t*-values reported in the text for the ROI analyses refer to regions identified in the color perception localizer task.

Region	X coor	Y coor	Z coor	Peak *t* Value
L Inferior Frontal Gyrus	−42	1	23	16.44
L Insula	−45	4	15	16.44
L Precuneus	−27	−62	36	16.19
R Middle Occipital Gyrus	21	−96	5	15.17
R Lingual Gyrus	9	−93	0	14.14
L Fusiform Gyrus	−45	−68	−14	14.04
L Cingulate Gyrus	−3	16	37	13.80
R Cingulate Gyrus	3	11	40	13.79
L Thalamus	−9	−20	4	13.68
L Inferior Occipital Gyrus	−27	−85	−13	13.20
L Lingual Gyrus	−9	−93	0	12.97
L Putamen	−21	0	−3	12.95
L Precentral Gyrus	−36	−6	56	12.51
R Fusiform Gyrus	21	−88	−11	12.25
L Inferior Parietal Lobule	−42	−30	39	12.07
L Cuneus	−24	−96	0	11.90
R Parahippocampal Gyrus	21	−32	−3	11.32
R Inferior Frontal Gyrus	33	20	−4	11.05
R Putamen	21	6	0	10.92
L Parahippocampal Gyrus	−24	−29	−4	10.87
L Supramarginal Gyrus	−36	−42	37	10.74
L Superior Frontal Gyrus	−18	−8	66	10.35
R Cuneus	12	−75	9	10.28
R Thalamus	12	−17	1	9.88
L Posterior Cingulate	−9	−31	22	9.76
L Postcentral Gyrus	−39	−23	59	9.48
R Inferior Occipital Gyrus	39	−96	−5	8.66
R Precentral Gyrus	36	−12	61	8.48
R Hippocampus	33	−44	2	7.39
R Superior Frontal Gyrus	21	−8	66	7.34
L Uncus	−33	−13	−32	7.08
R Precuneus	27	−65	31	7.07
R Caudate (tail)	30	−43	12	6.98