Simple line drawings suffice for functional MRI decoding of natural scene categories

Dirk B. Walther[a,1], Barry Chai[b], Eamon Caddigan[c,d], Diane M. Beck[c,d], and Li Fei-Fei[b]

[a]Department of Psychology, The Ohio State University, Columbus, OH 43210; [b]Department of Computer Science, Stanford University, Stanford, CA 94305; [c]Beckman Institute, University of Illinois at Urbana–Champaign, Urbana, IL 61801; and [d]Department of Psychology, University of Illinois at Urbana–Champaign, Champaign, IL 61820

Edited by Anne Treisman, Princeton University, Princeton, NJ, and approved April 22, 2011 (received for review October 22, 2010)

Humans are remarkably efficient at categorizing natural scenes. In fact, scene categories can be decoded from functional MRI (fMRI) data throughout the ventral visual cortex, including the primary visual cortex, the parahippocampal place area (PPA), and the retrosplenial cortex (RSC). Here we ask whether, and where, we can still decode scene category if we reduce the scenes to mere lines. We collected fMRI data while participants viewed photographs and line drawings of beaches, city streets, forests, highways, mountains, and offices. Despite the marked difference in scene statistics, we were able to decode scene category from fMRI data for line drawings just as well as from activity for color photographs, in primary visual cortex through PPA and RSC. Even more remarkably, in PPA and RSC, error patterns for decoding from line drawings were very similar to those from color photographs. These data suggest that, in these regions, the information used to distinguish scene category is similar for line drawings and photographs. To determine the relative contributions of local and global structure to the human ability to categorize scenes, we selectively removed long or short contours from the line drawings. In a category-matching task, participants performed significantly worse when long contours were removed than when short contours were removed. We conclude that global scene structure, which is preserved in line drawings, plays an integral part in representing scene categories.

scene perception | line art | multivoxel pattern analysis | neuroimaging | visual processing

Humans have captured scenes of everyday life with simple lines since prehistoric times (1). Line drawings pervade the history of art in most cultures on Earth (see Fig. 1 *A*–*C* for examples). Although line drawings lack many of the defining characteristics seen in the real world (color, most texture, most shading, etc.), they nevertheless appear to capture some essential structure that makes them useful as a way to depict the world for artistic expression or as a visual record. In fact, children use "boundary lines" or "embracing lines" to define the shapes of objects and object parts in their first attempts to depict the world around them (2) (see Fig. 1*D* for an example). The natural ability of humans to recognize and interpret line drawings has also made them a useful tool for studying objects and scenes (3–5).

In the experiments described in this article, we used line drawings of natural scenes of six categories (beaches, city streets, forests, highways, mountains, and offices; Fig. S1) to explore the human ability to efficiently categorize natural scenes. Humans can recognize the gist of a scene with presentations as short as 120 ms (6), even when their attention is engaged elsewhere in the visual field (7, 8). This gist can include many details beyond a basic category-level description (9–11).

One reason why humans may be so fast at processing natural scenes is that our visual system evolved to efficiently encode statistical regularities in our environment. However, we can nevertheless recognize and categorize line drawings of natural scenes despite their having very different statistical properties from photographs (Fig. 2*D* and Fig. S1). Indeed, the fact that early artists as well as young children represent their world with line drawings suggests that such depictions capture the essence of our natural world. How are line drawings processed in the brain? What are the similarities and the differences in processing line drawings and color photographs of natural scenes? Here we approach these questions by decoding scene category information from patterns of brain activity measured with functional MRI (fMRI) in observers viewing photographs and line drawings of natural scenes.

We have previously found that information about scene category is contained in patterns of fMRI activity in the parahippocampal place area (PPA), the retrosplenial cortex (RSC), the lateral occipital complex (LOC), and the primary visual cortex (V1) (12). PPA activity patterns appear to be linked most closely to human behavior (12), and activity in V1, PPA, and RSC elicited by good exemplars of scene categories contains significantly more scene category–specific information than patterns elicited by bad exemplars do (13). Interestingly, inspection of the average images of good and bad exemplars of a category suggests that good exemplars may contain more defined global structure apparent in the images than bad exemplars (13), suggesting that features that capture global information in the image may play a particularly important role in scene categorization.

In the work presented here, we examine the effect of scene structure, or layout, more directly by stripping down images of natural scenes to the bare minimum, mere lines. If such structure really is critical, does this mean that we can decode scene categories from the brain activity elicited by line drawings? If so, how are these representations related to those from full-color photographs?

Finally, what aspects of the structure preserved in line drawings allow us to categorize natural scenes? With color and most texture out of the picture, all that is left is the structure of the scene captured by the lines. In a behavioral experiment, we attempt to discriminate the contributions of local versus global structure by selectively removing long or short contours from line drawings of natural scenes.

Results

fMRI Decoding. To investigate and compare the neural activation patterns elicited by color photographs and line drawings of natural scenes, we asked participants to passively view blocks of images while inside an MRI scanner. In half of the runs, participants saw color photographs of six categories (beaches, city streets, forests, highways, mountains, and offices) and, in the other half, viewed line drawings of the same images (Fig. 2 and Fig. S1). The order of blocks with photographs and corresponding line drawings was counterbalanced across participants. Blood oxygen level–dependent activity was recorded in 35 coronal slices, covering approximately the posterior two-thirds of the brain.

We analyzed blood oxygen level–dependent activity in retinotopic areas in the visual cortex as well as in the PPA, RSC, and LOC. V1 is an interesting brain region for comparing photographs with line drawings because it is optimized for extracting

Author contributions: D.B.W., B.C., E.C., D.M.B., and L.F.-F. designed research; D.B.W., B.C., and E.C. performed research; B.C. contributed new reagents/analytic tools; D.B.W. analyzed data; and D.B.W., D.M.B., and L.F.-F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: bernhardt-walther.1@osu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1015666108/-/DCSupplemental.

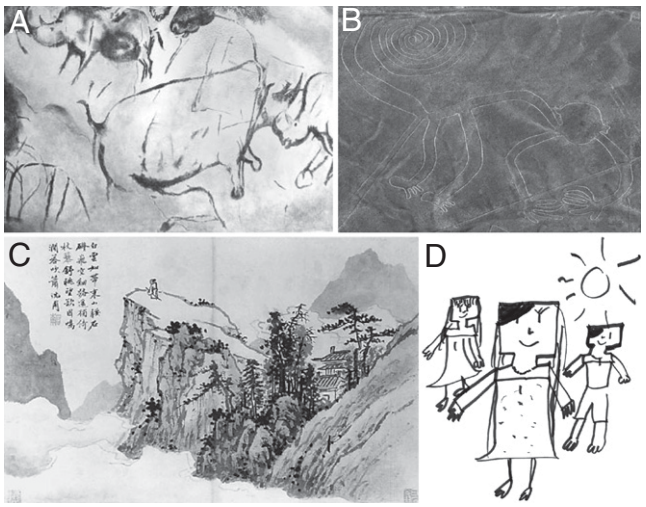

Fig. 1. Examples of line art. (*A*) Cave painting at Chauvet, France, *ca.* 30,000 B.C. (*B*) Aerial photograph of the picture of a monkey as part of the Nazca Lines geoglyphs, Peru, *ca.* 700–200 B.C. (*C*) Shen Zhou (A.D. 1427–1509): Poet on a mountain top, ink on paper, China (The Nelson-Atkins Museum of Art, Kansas City, MO). (*D*) Line drawing by 7-y-old I. Lleras (A.D. 2010).

edges and lines from the retinal image (14, 15). Therefore, it is conceivable that V1 may represent photographs and line drawings in a similar way. Visual areas V2, VP (known as V3v, for ventral V3, in some nomenclatures), and V4 are of interest because they build more complex representations based on the information in V1, including representations that rely on more global aspects of the image (16–18). The PPA and the RSC, which have been shown to prefer scenes over objects and other visual stimuli (19, 20), are of interest in this analysis because they (and, to some extent, the LOC) contain information about scene category in their activity patterns that are closely linked to behavioral scene-categorization performance (12).

A separate fMRI session for measuring retinotopy allowed us to delineate areas V1 and V4, but the border between V2 and VP was not apparent in all participants, so we considered those regions as a group (V2+VP). The PPA, RSC, and LOC were defined based on linear contrasts in a standard localizer session showing faces, scenes, objects, and scrambled objects (*SI Materials and Methods*).

Decoding Photographs. Using only data from the runs when participants viewed color photographs, we trained and tested a decoder to predict which of the six scene categories participants were viewing in a leave-one-run-out (LORO) cross-validation procedure with the data from each of the regions of interest (ROIs) in turn. We found above-chance (>16.7%) decoding accuracy all along the visual-processing hierarchy (Fig. 3): 24% in V1 [$t(9) = 3.03$, $P = 0.0071$; one-tailed t test versus chance level], 27% in V2+VP [$t(9) = 7.40$, $P = 2.05 \times 10^{-5}$], 24% in V4 [$t(9) = 5.18$, $P = 2.9 \times 10^{-4}$], 32% in the PPA [$t(9) = 3.97$, $P = 0.0016$], and 23% in the RSC [$t(9) = 3.02$, $P = 0.0073$]. These results confirm our earlier observations that these regions contain information that distinguishes among natural scene categories (12). Unlike with our previous data (12), decoding accuracy of 21% in the LOC did not reach significance [$t(9) = 1.5$, $P = 0.080$], possibly because of differences in the scene categories used in the two studies and, in particular, differences in the number of objects that uniquely identify a category. For example, cars appeared in both highways and city streets in the current experiment, whereas cars uniquely identified highways in the previous experiment. Although the LOC has been shown to encode objects in scenes (21) as well as the spatial relationship between objects (22, 23), its contribution to the categorization of scenes likely depends on such diagnostic objects.

Decoding Line Drawings. How does decoding fare when the photographs are reduced to mere lines? To address this question, we repeated the LORO cross-validation analysis using the runs during which participants saw line drawings of natural scenes. Decoding of category from line drawings was not only possible in all of the same brain regions as was decoding from color photographs [V1: 29%, $t(9) = 4.71$, $P = 5.5 \times 10^{-4}$; V2+VP: 27%, $t(9) = 4.09$, $P = 0.0014$; V4: 26%, $t(9) = 3.72$, $P = 0.0024$; PPA: 29%, $t(9) = 7.24$, $P = 2.4 \times 10^{-5}$; and RSC: 23%, $t(9) = 3.17$, $P = 0.0057$; all one-tailed t tests versus chance level], but, even more surprisingly, decoding accuracy for line drawings was also at the same level as decoding for color photographs in all ROIs [V1: $t(9) = 1.18$, $P = 0.27$; V2+VP: $t(9) = 0.15$, $P = 0.88$; V4: $t(9) = 0.77$, $P = 0.46$; PPA: $t(9) = 0.79$, $P = 0.45$; and RSC: $t(9) = 0.02$, $P = 0.98$; all paired, two-tailed t tests] (Fig. 3). That is, despite marked differences in scene statistics and considerable degradation of information, line drawings can be decoded as accurately as photographs in every region tested. In fact, in V1, we even saw somewhat higher, albeit not significantly, decoding accuracy for line drawings than for photographs. Although we predicted that line drawings might contain some essential features of natural scene categories, thus making decoding possible, it is surprising that there appears to be no further benefit for photographs. One should keep in mind, however, that the line drawings used in this study were created by trained artists tracing just those contours in the image that best captured the scene. The clearly defined contours may make it easier to extract, in V1 in particular, differences and regularities in orientations in certain image regions across categories. Such prevalent orientation in-

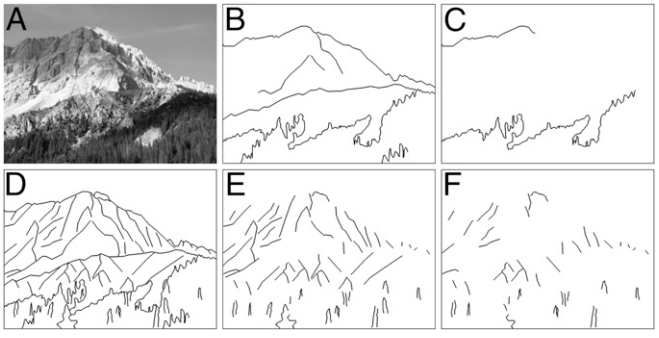

Fig. 2. Example of a photograph of a mountain scene (*A*) and the corresponding line drawing (*D*). Also shown are the degraded line drawings with 50% (*B* and *E*) or 75% (*C* and *F*) of the pixels of the line drawings removed through deletion of the shortest (*B* and *C*) or the longest (*E* and *F*) contours. See Fig. S1 for examples for all six categories.

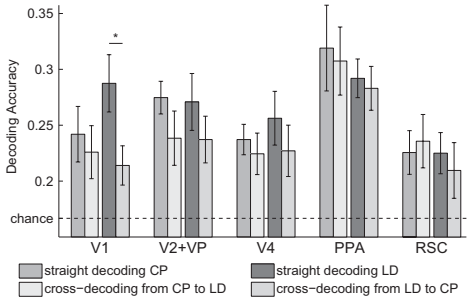

Fig. 3. Accuracy of straight decoding of scene categories from brain activity while participants viewed color photographs (CP) and line drawings (LD) as well as cross-decoding from CP to LD and from LD to CP. Accuracy was significantly above chance (1/6) in all conditions, except for cross-decoding from LD to CP in the RSC. Straight decoding for CP and LD was statistically the same for all ROIs. The drop from straight decoding LD to cross-decoding from LD to CP was significant in V1 (*$P < 0.05$). Error bars are SEM over 10 participants.

formation has previously been shown to be decodable from multivoxel activity patterns in V1 (24).

In the fMRI experiment, blocks of photographs were alternated with blocks of line drawings. Could the high decoding accuracy for line drawings be because of the visual imagery of the corresponding color photographs presented in the preceding block? To address this possibility, we compared decoding rates for the five participants who saw photographs first to the five participants who saw line drawings first. In no ROI did we see a significant improvement in the decoding of line drawings for the group that saw the photographs first ($P > 0.22$; one-tailed, unpaired t tests), and, in fact, in all regions but V1, the decoding rates were numerically lower for the photographs-first group.

The similar decoding accuracy for line drawings and photographs raises an interesting question: Are we decoding similar information in the two cases? Although the representations of color photographs and line drawings must differ to some extent, it may be that those features that denote scene category are best captured by the edges and lines in the image. However, before making such a conclusion, we must consider an alternative possibility: that the brain uses different, but equally effective, information to derive scene category for photographs and line drawings.

Decoding Across Image Types. If the brain uses different information to decode scene category from photographs and line drawings, then we should see a decrease in decoding accuracy when we train on one image type and test on the other (e.g., train on line drawings and test on photographs). If, on the other hand, the information that distinguishes between categories is similar for photographs and line drawings, then we should see similar decoding accuracies for cross-decoding from one image type to the other as we do for straight decoding from one image type to the same image type.

To test these two hypotheses, we crossed the factors of training on photographs or line drawings with testing on either the same image type (straight decoding) or the other image type (cross-decoding) (Fig. 3). For each ROI, we performed a two-way ANOVA with two independent variables: (*i*) training on photographs versus line drawings and (*ii*) straight decoding versus cross-decoding. Neither the main effects of training and straight versus cross-decoding nor their interaction reached significance in any ROI, although there was a marginal drop in accuracy for cross-decoding in V1 ($F_{1,9} = 3.74$, $P = 0.061$). These data suggest, with the possible exception of V1, that decoding category from line drawings relies on similar information as decoding category from photographs does. We cannot, however, draw any strong conclusions from null results. We therefore tested the similarity of category representations elicited by line drawings and photographs further by turning to decoding errors.

Correlating Decoding Errors. As an additional measure of the similarity of the activity patterns elicited by photographs and line drawings, we compared decoding errors from the two image types. The particular errors, or confusions, made by the decoder, which are recorded in a confusion matrix, can reveal aspects of the underlying representation of our categories (12, 25). If line drawings and photographs evoke similar category representations in an area, then the confusions made by the decoder are bound to be similar. The rows of a confusion matrix indicate which scene category was presented to the participants, and the columns represent the prediction of the decoder. Diagonal elements correspond to correct decoding, and off-diagonal elements correspond to decoding errors (Fig. 4). For this analysis, we combined error entries for symmetric pairs of categories (e.g., confusing beaches for highways was combined with confusing highways for beaches), and we averaged confusion matrices over all 10 participants. To compare the confusions elicited by photographs with those elicited by line drawings, we computed the Pearson correlation coefficient for correlating the off-diagonal elements of the confusion matrix for the two types of images (Fig. S2).

Error patterns when decoding from line drawings were highly correlated with the error patterns when decoding from photographs in the PPA [$r(13) = 0.86$, $P = 3.3 \times 10^{-5}$] and the RSC [$r(13) = 0.76$, $P = 9.7 \times 10^{-4}$], somewhat correlated in V4 [$r(13) = 0.52$, $P = 0.047$], but not correlated in V1 [$r(13) = 0.47$, $P = 0.075$] or V2+VP [$r(13) = 0.47$, $P = 0.081$]. These results lend further support to the claim that the category-specific information elicited by color photographs or line drawings is most similar in the PPA, somewhat similar in V4, and least similar in V1 and V2+VP. In other words, this pattern of results suggests an increase in the similarity of the representations of photographs and line drawings as we ascend the visual hierarchy.

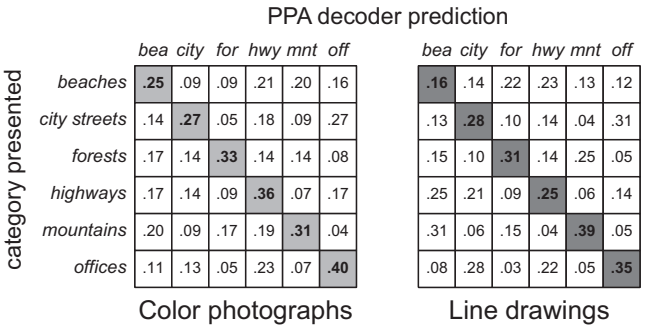

Fig. 4. Confusion matrices for straight decoding of scene categories from color photographs and line drawings in the PPA. Diagonal entries are correct decoding rates for the respective categories, and off-diagonal entries indicate decoding errors.

To rule out the possibility that the differences across brain regions could be attributed to differences in the number of voxels in each ROI, we repeated the analysis by subsampling equal numbers of voxels in each ROI. Although slightly different numerically, the results obtained from this analysis confirm the same patterns of effects reported above (*SI Materials and Methods*). Finally, we conducted a whole-brain searchlight analysis to look for areas beyond our predefined ROIs that encode scene category. In addition to the known ROIs, we found a small cluster of voxels with decoding accuracy significantly above chance in the left precuneus (Figs. S3 and S4).

Global Versus Local Structure. Since the information contained within line drawings is sufficient for categorization and leads to comparable decoding performance as with photographs, we can begin to ask what about the line drawings makes such surprisingly good performance possible. In their well-accepted model, Oliva and Torralba used properties of the amplitude spectrum of spatial frequencies (the "spatial envelope") to discriminate categories of natural scenes (26). However, the spatial-frequency spectra of line drawings are radically different from those of the corresponding photographs (Fig. S5), yet humans can still categorize them with ease. What properties of line drawings allow humans to categorize them like this? With shape or structure being the main image property retained in the line drawings, we here ask whether global or local structure is an important factor for categorization.

What do we mean by global and local structure in this context? Our line drawings were generated by trained artists, who traced the contours of the color photographs on a graphics tablet. We consider all lines drawn in one stroke without lifting the pen from the tablet as belonging to the same contour. Each contour consists of a series of straight line segments, with the end point of one line segment being the starting point of the next. We hypothesized that long contours are more likely to reflect global structure and short contours are more likely to represent local structure in the image. Here we interpret "global structure" as representing gross areas in the image, such as sky, water, sand, or building façade, as opposed to smaller details, such as leaves, windows in buildings, or individual ridges in a mountainside. Intuitively, such large image areas are likely to be separated by long contours, whereas short contours are more likely to delineate the smaller details. How can we test whether this pre-

sumed connection between line length and global/local structure holds in our images? If a contour separates two large areas of the image, then the areas on either side of the contour should differ in their image properties, e.g., in their color distributions. If a contour only defines local structure within a large area, then the color distributions on either side are more likely to be similar.

To assess the regions separated by our contours, we determined the color histogram distance (CHD) between regions that fell on either side of the line segments in the original color photographs (Fig. S6). Within the area on one side of a line segment, we divided each of the three color channels (red, green, and blue) into four equally sized bins ($4^3 = 64$ bins in total), counted the numbers of pixels in each bin, and divided by the total number of pixels in the area. This procedure was repeated for the image area on the other side of the line segment. We then computed the CHD as the Euclidean distance between the two histograms and averaged the CHD over all line segments belonging to the same contour.

If long contours do indeed separate global regions of the image, then we should expect that the image areas on opposite sides of the contour differ from each other more for long than for short contours, i.e., that their CHDs are larger. For each image, we sorted the contours by their total length and computed the average CHD separately for the top and bottom quartile of contours, as measured by the fraction of pixels covered. The average CHD for long contours was significantly larger than that for short contours [long: 0.589; short: 0.555; $t(474) = 8.16$, $P = 3.01 \times 10^{-15}$; paired, two-tailed t test; for CHDs for the individual categories see Table S2]. These data suggest that long contours do indeed separate global structures in the image, defined here as large regions in the image that differ in their image properties (captured here by the color histogram), whereas short contours are less likely to do so.

Degrading Line Drawings. Now we can investigate the role of global versus local structure in the ability to categorize scenes by systematically removing contours from the line drawings (omitting them when rendering the line drawings), starting either with the longest or the shortest contours. Which deletion impacts human categorization performance more? Our procedure for modifying the line drawings also allowed us to ask how far we could push the degradation of the line drawings while leaving categorization performance intact. In particular, we deleted contours from our line drawings such that 50% and even 75% of the pixels were removed (Fig. 2). Removing contours also led to the removal of angles and intersections from the line drawings: when removing 50% of the pixels starting from the shortest/longest contours, on average 61% (SD = 12%)/44% (SD = 10%) of the angles and line intersections were removed; when removing 75% of the pixels, 81% (SD = 9%)/70% (SD = 8%) of the angles were removed. In other words, fewer angles and intersections were removed when removing long contours.

We used a same/different task to assess categorization performance for the different forms of line degradation behaviorally. On each trial, participants were shown three line drawings in rapid succession (Fig. 5*A*). The first and the third image were line drawings of two different natural scenes from the same category (e.g., both city streets). Participants were asked to indicate whether the second image was a line drawing of that same category or of a different one by pressing "S" (same) or "D" (different) on the computer keyboard. Chance level was at 50%. To effectively mask the critical second image, the first and third image were shown as white line drawings on a black background, and the second image was shown as black on white. We chose this same/different judgment task, because (*i*) the same/different judgment was easier than a six-alternative forced-choice category judgment, and (*ii*) we found other line drawings with reversed contrast to be the most effective forward and backward masks for line drawings, allowing us to adjust the difficulty of the task for each participant by modifying the presentation duration of the images.

After a practice phase to familiarize participants with the experiment, the presentation duration of each of the three images was systematically reduced from 300 ms in a staircasing procedure. Staircasing was terminated when a stable performance level of 70% was reached (SD over one block less than screen refresh rate). Remarkably, this level of accuracy in the same/different judgment was still possible with presentation times as short as 102 ms per image (mean: 175 ms; SD: 66 ms). This finding demonstrates that categorizing line drawings of natural scenes can be achieved with a single brief glance.

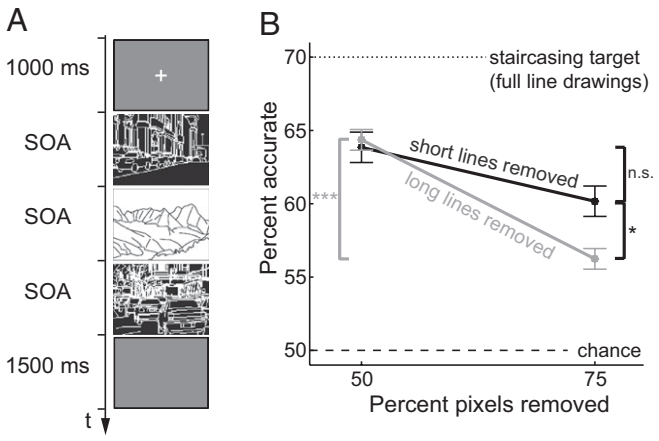

Fig. 5. (*A*) Time course of a single trial of the behavioral experiment. Stimulus onset asynchrony (SOA) was adjusted for each participant by using a staircasing procedure with intact line drawings. (*B*) Performance in the behavioral experiment for line drawings with long or short contours removed such that 50% or 75% of the pixels were deleted. Performance is significantly ($P < 0.05$) above chance in all cases, and, for 75%, it is significantly better for line drawings with short than long contours removed. Error bars are SEM over 13 participants. *$P < 0.05$, ***$P < 0.001$.

In the subsequent test phase, presentation times remained constant at the values obtained for each participant during staircasing. To test the effect of line deletion on behavioral performance, we substituted four different versions of degraded images for the second image, omitting lines starting from the longest/shortest contours until 50% or 75% of the pixels were removed (see Fig. 2 and Fig. S1 for examples). These four types of degradations were counterbalanced over target categories and over same/different trials for each participant.

Remarkably, performance for all four types of degradation was still significantly above chance ($P < 0.05$; Fig. 5*B*). Performance was significantly better for line drawings with 50% of the pixels removed than with 75% removed when those pixels were removed in the form of long contours [$t(12) = 5.76$, $P = 9.0 \times 10^{-5}$] but not when they were removed in the form of short contours [$t(12) = 1.77$, $P = 0.10$; paired, two-tailed t tests]. In other words, degrading line drawings beyond 50% hurt performance only when long contours were deleted. Note that this result is unlikely to be caused by differences in the number of deleted angles and intersections because fewer angles and intersections were removed when removing long rather than short contours.

Interestingly, performance with 50% degradation was the same, no matter which type of contours was removed, presumably because there was still sufficient scene information present in both cases. However, when the images were degraded further, a significant difference between the long and short lines emerged: performance with 75% degradation was significantly lower when removing long rather than short contours [$t(12) = 2.67$, $P = 0.020$], although the fraction of pixels that remained in the image was identical. This result means that long contours, which we argue capture global image information better than short contours, were more important for correctly identifying scene categories. Together, these results suggest a primary role for global rather than local image information in defining scene category. Global image information may be an important factor in computing scene layout, a role often ascribed to the PPA (19),

which has been heavily implicated in natural scene categorization here and elsewhere (12, 27, 28).

To test how these behavioral results relate to the results from decoding scene category from fMRI activity, we compared the patterns of errors in the two experiments (*SI Materials and Methods* and Fig. S7). The high error correlation with behavior in the PPA ($r = 0.69$, $P = 0.0043$), first established for a six-alternative forced-choice task in ref. 12, strengthens the close link between this area's neural activity and human scene-categorization behavior for the images and categories themselves as opposed to the task being performed.

Discussion

Artistic expression by means of line drawings pervades most human cultures. Our results show that, despite being drastically reduced versions of natural scenes, line drawings capture the structure of scenes that enables scene categorization. The neural activity elicited by line drawings contains sufficient information to allow for six-way decoding of scene categories along the visual-processing hierarchy, starting from V1 through V4 all the way up to the PPA and RSC. Remarkably, decoding scene categories from the neural activity patterns elicited by line drawings was just as accurate as decoding from activity generated by color photographs.

Successful cross-decoding as well as similar decoding error patterns between color photographs and line drawings suggest that the two image types generate similar category-specific activity. We saw an increase in the correlation of error patterns along the visual hierarchy, starting from low correlation in V1 and V2+VP, through intermediate but significant correlation in V4, to high correlation in the PPA and RSC. We conclude that the structural information preserved in line drawings is sufficient to generate a strong category-specific representation of natural scene categories that is highly compatible with the one generated by color photographs in the PPA and RSC but is less compatible in early visual areas.

In V1, accuracy of cross-decoding from line drawings to photographs was significantly lower than straight decoding of line drawings. Furthermore, the error patterns recorded in V1 for photographs and line drawings were not correlated significantly, which is compatible with the view that early visual areas allow for the decoding of natural scene categories based on common low-level features among the exemplars of a given scene category but that these features differ to some extent between photographs and line drawings. This difference in the robustness of the category representations between V1 and later areas is also compatible with the view that, although scene category may be distinguishable on the basis of low-level V1-like features, these features are not necessarily all used by humans to make such categorizations (12).

Compared with a fully textured and colored photograph, the line drawing of an image is a drastically impoverished version of the picture. Moreover, converting photographs to line drawings alters the spatial-frequency spectrum considerably. However, our behavioral as well as our decoding experiments show that the human ability to categorize natural scenes is robust to the deletion of this information. These results argue against the rich statistical properties of natural scenes as being necessary for quick and accurate scene categorization.

Of course, the brain may still use color, shading, and texture information in photographs to aid in categorizing a scene, but our work suggests that the features preserved in line drawings are not only sufficient for scene categorization but that they are also likely used in categorizing both line drawings and photographs. What could those features be?

We hypothesized that scene structure, which is preserved in line drawings, is important for categorization. Specifically, we predicted that global structure, which we defined as contours that separate large but relatively homogenous regions of the image, would be more important than local structure, because the PPA has been implicated in both scene layout (19) and scene categorization (12, 28) and because good category exemplars appear to have more consistent global scene structure (13). To test this hypothesis, we modified line drawings by selectively removing long or short contours. We found that accuracy of a same/different category judgment was significantly lower when we removed 75% of the pixels from long contours than when we removed the same percentage of pixels from short contours. Thus, it appears that global structure (better captured by long contours) is more important for categorizing scenes from line drawings than local structure (captured in short contours).

In summary, we have found that we can perform six-way decoding of scene category equally well from brain activity generated from line drawings and color photographs in early to intermediate visual areas (V1, V2+VP, and V4) as well as in the PPA and RSC. The brain activity in the PPA and RSC for the two kinds of images is particularly compatible, as shown by cross-decoding and the analysis of decoding errors. By systematically degrading the line drawings, we have determined that global structure is likely to play a more prominent role than local structure in determining scene categories from line drawings.

Materials and Methods

Participants. Ten volunteers (mean age 29.5 y, SD 4.9 y; 6 female) participated in the fMRI experiment, and 13 volunteers (mean age 21.6 y, SD 4.3 y; 11 female) participated in the behavioral experiment. Both experiments were approved by the Institutional Review Board of the University of Illinois. All participants were in good health with no past history of psychiatric or neurological diseases and gave their written informed consent. Participants had normal or corrected-to-normal vision.

Images. Color photographs (CPs) of six categories (beaches, city streets, forests, highways, mountains, and offices) were downloaded from the Worldwide Web via multiple search engines. Workers at the Amazon Mechanical Turk web service were paid to rate the quality of the images as exemplars of their respective category. For the experiments in this article, we only used the 80 highest-rated images for each category. Images with fire or other potentially upsetting content were removed from the experiment (five images total). See Torralbo et al. (13) for details of the image-ratings procedure. All photographs were resized to a resolution of 800 × 600 pixels for the experiment.

Line drawings (LDs) of the photographs were produced by trained artists at the Lotus Hill Research Institute by tracing contours in the color photographs via a custom graphical user interface. The order and coordinates of all line strokes were recorded digitally to allow for later reconstruction of the line drawings at any resolution. For the experiments, line drawings were rendered at a resolution of 800 × 600 pixels by drawing black lines on a white background.

fMRI Experiment. In the fMRI experiments, participants passively viewed blocks of eight images of the same category. Each image was presented for 2 s without a gap. A 12-s fixation interval was inserted before the first block of each run, between blocks, and after the last block. Each run contained six blocks, one for each of the six categories. CPs and LDs were presented in alternating runs, with the order of the image type counterbalanced across participants. The order of blocks within runs as well as the selection and order of images within blocks were randomized for all odd-numbered runs. In even-numbered runs, the block structure of the preceding odd-numbered run was repeated, except that the other image type (CP or LD) was used. Two participants saw 14 runs, the other eight participants saw 16 runs total. Images were presented with a back-projection system at a resolution of 800 × 600 pixels, corresponding to a visual angle of 23.5° × 17.6°.

fMRI Data Acquisition. MRI images were recorded on a 3-T Siemens Allegra. Functional images were obtained with a gradient echo, echo-planar sequence (repetition time, 2 s; echo time, 30 ms; flip angle, 90°; matrix, 64 × 64 voxels; field of view, 22 cm; 35 coronal slices, 2.8-mm thick with 0.7-mm gap; in-plane resolution, 2.8 × 2.8 mm). We also collected a high-resolution (1.25 × 1.25 × 1.25 mm voxels) structural scan (MPRAGE; repetition time, 2 s; echo time, 2.22 ms, flip angle, 8°) in each scanning session to assist in registering our echo planar images across sessions.

fMRI Data Analysis. Functional data were registered to a reference volume (45th volume of the eighth run) by using AFNI (29) to compensate for subject motion during the experiment. Data were then normalized to represent percentage signal change with respect to the temporal mean of each run. No other smoothing or normalization steps were performed. Brain volumes corresponding to the blocks of images in each run were extracted from the

time series with a lag of 4 s to account for the delay in the hemodynamic response. Data were processed separately for each of the ROIs.

Activation data from all except one of the runs with color photographs were used in conjunction with the corresponding scene category labels to train a support vector machine (SVM) classifier (linear kernel, $C = 0.02$) using LIBSVM. Presented with the data from the left-out run, the SVM classifier generated a prediction for the scene category labels for each brain acquisition. Disagreements in the predicted labels for the eight brain volumes belonging to the same block were resolved by majority voting, resulting in the label predicted most frequently among the eight volumes to be chosen as the label for the entire block. Ties were broken by adopting the label with the highest decision value in the SVM classifier. By repeating this procedure such that each of the CP runs was left out once (LORO cross-validation), predictions for the scene categories were generated for the blocks in each run (straight decoding). Decoding accuracy was computed as the fraction of correct predictions over all runs. The same LORO cross-validation procedure was used to determine decoding accuracy for the runs with line drawings.

To see whether category-specific information learned from activation patterns elicited by color photographs generalizes to line drawings, we modified the analysis procedure slightly. As before, the SVM classifier was trained on the fMRI activation data from all except one of the CP runs. Instead of testing it with the data from the left-out CP run, however, the classifier was tested with the fMRI data from the LD run whose block and image structure was the same as that of the left-out CP run (cross-decoding). The analysis was repeated such that each CP run was left out once, thus generating predictions for each of the LD runs. Decoding accuracy was again assessed as the fraction of correct predictions of the scene category label. Cross-decoding from line drawings to color photographs was computed in an analogous manner, training the classifier on all except one of the LD runs and testing it on the CP run corresponding to the left-out LD run.

Behavioral Experiment. In the behavioral experiment, participants were asked to perform a same/different category judgment among three images presented in rapid serial visual presentation. Stimuli were presented on a CRT monitor at a resolution of 800 × 600 pixels, subtending 23° × 18° of visual angle. At the beginning of each trial, a white fixation cross was presented at the center of the screen on a 50% gray background for 1,000 ms. Then a sequence of three images was shown at full screen size in rapid succession for a predetermined duration, without gap, followed by 1,500 ms of blank screen to allow participants to respond (Fig. 5*A*). The first and the third image were line drawings of two different natural scenes from the same category. In half of the trials, the second image was yet another line drawing from the same category, and in the other half, it was a line drawing from a different category. Participants were asked to press "S" on the computer keyboard when the second image was from the same category as the first and third or "D" when it was from a different category. To effectively mask the critical second image, the first and third image were shown as white line drawings on a black background, and the second image was shown as black on white.

The experiment was composed of three parts: practice, staircasing, and testing. During practice and staircasing, the second images were intact line drawings, and a tone was given as feedback for erroneous responses. The presentation time for the three images was systematically reduced from 700 ms to 300 ms during practice. In the subsequent staircasing phase, presentation time was reduced further by using the Quest staircasing algorithm (30) until a stable accuracy level of 70% was achieved. This presentation duration, determined individually for each participant, was then used in the testing phase. Average presentation time after staircasing was 175 ms (SD: 66 ms).

In the testing phase, the second images were degraded versions of the original line drawings. Contiguous sets of lines (contours) in each line drawing were sorted by their total length. Degraded versions of the line drawings were obtained by omitting contours such that 50% or 75% of the pixels were removed, starting from the longest or starting from the shortest contours (Fig. 2 and Fig. S1). When necessary, individual line segments were shortened or removed from contours to achieve the accurate percentage of pixels. These four types of degradations were counterbalanced over target category and over same/different trials for each participant. Accuracy was measured as the fraction of trials, in which participants correctly responded by pressing "S" for same trials and "D" for different trials.

ACKNOWLEDGMENTS. This work was funded by National Institutes of Health Grant 1 R01 EY019429 (to L.F.-F., D.M.B., and D.B.W.).

1. Clottes J (2000) *Chauvet Cave* (*ca. 30,000 B.C.*). *Heilbrunn Timeline of Art History* (The Metropolitan Museum of Art, New York, NY), Vol 2010.
2. Goodnow J (1977) *Children's Drawing* (Fontana/Open Books, London).
3. Biederman I, Ju G (1988) Surface versus edge-based determinants of visual recognition. *Cognit Psychol* 20:38–64.
4. Biederman I, Mezzanotte RJ, Rabinowitz JC (1982) Scene perception: Detecting and judging objects undergoing relational violations. *Cognit Psychol* 14:143–177.
5. Snodgrass JG, Vanderwart M (1980) A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *J Exp Psychol Hum Learn* 6:174–215.
6. Potter MC (1976) Short-term conceptual memory for pictures. *J Exp Psychol Hum Learn* 2:509–522.
7. Fei-Fei L, VanRullen R, Koch C, Perona P (2005) Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Vis Cogn* 12:893–924.
8. Li FF, VanRullen R, Koch C, Perona P (2002) Rapid natural scene categorization in the near absence of attention. *Proc Natl Acad Sci USA* 99:9596–9601.
9. Fei-Fei L, Iyer A, Koch C, Perona P (2007) What do we perceive in a glance of a real-world scene? *J Vis* 7:10.
10. Greene MR, Oliva A (2009) The briefest of glances: The time course of natural scene understanding. *Psychol Sci* 20:464–472.
11. Wolfe JM (1998) Visual memory: What do you know about what you saw? *Curr Biol* 8: R303–R304.
12. Walther DB, Caddigan E, Fei-Fei L, Beck DM (2009) Natural scene categories revealed in distributed patterns of activity in the human brain. *J Neurosci* 29:10573–10581.
13. Torralbo A, et al. (2009) Categorization of good and bad examples of natural scene categories. *J Vis*, 10.1167/9.8.940.
14. De Valois RL, De Valois KK (1980) Spatial vision. *Annu Rev Psychol* 31:309–341.
15. Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160:106–154.
16. Gallant JL, Connor CE, Rakshit S, Lewis JW, Van Essen DC (1996) Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J Neurophysiol* 76:2718–2739.
17. Pasupathy A, Connor CE (1999) Responses to contour features in macaque area V4. *J Neurophysiol* 82:2490–2502.
18. Peterhans E, von der Heydt R (1993) Functional organization of area V2 in the alert macaque. *Eur J Neurosci* 5:509–524.
19. Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598–601.
20. O'Craven KM, Kanwisher N (2000) Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J Cogn Neurosci* 12:1013–1023.
21. Peelen MV, Fei-Fei L, Kastner S (2009) Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature* 460:94–97.
22. Kim JG, Biederman I (2010) Where do objects become scenes? *Cereb Cortex*, 10.1093/cercor/bhq240.
23. Macevoy SP, Epstein RA (2009) Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr Biol* 19:943–947.
24. Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8:679–685.
25. Walther DB, Beck DM, Fei-Fei L (2011) To err is human: Correlating fMRI decoding and behavioral errors to probe the neural representation of natural scene categories. *Understanding Visual Population Codes—Toward a Common Multivariate Framework for Cell Recording and Functional Imaging*, eds Kriegeskorte N, Kreiman G (MIT Press, Cambridge, MA).
26. Oliva A, Torralba A (2001) Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int J Comput Vis* 42:145–175.
27. Bar M, Aminoff E (2003) Cortical analysis of visual context. *Neuron* 38:347–358.
28. Epstein RA, Higgins JS (2007) Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cereb Cortex* 17:1680–1693.
29. Cox RW (1996) AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173.
30. King-Smith PE, Grigsby SS, Vingrys AJ, Benes SC, Supowit A (1994) Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation. *Vision Res* 34:885–912.