Behavioral and Neuroimaging Evidence for a Contribution of Color and Texture Information to Scene Classification in a Patient with Visual Form Agnosia

Jennifer K. E. Steeves[1], G. Keith Humphrey[1], Jody C. Culham[1], Ravi S. Menon[2], A. David Milner[3], and Melvyn A. Goodale[1]

Abstract

■ A common notion is that object perception is a necessary precursor to scene perception. Behavioral evidence suggests, however, that scene perception can operate independently of object perception. Further, neuroimaging has revealed a specialized human cortical area for viewing scenes that is anatomically distinct from areas activated by viewing objects. Here we show that an individual with visual form agnosia, D.F., who has a profound deficit in object recognition but spared color and visual texture perception, could still classify scenes and that she was fastest when the scenes were presented in the appropriate color. When scenes were presented as black-and-white images, she made a large number of errors in classification. Functional magnetic resonance imaging revealed selective activation in the parahippocampal place area (PPA) when D.F. viewed scenes. Unlike control observers, D.F. demonstrated higher activation in the PPA for scenes presented in the appropriate color than for black-and-white versions. The results demonstrate that an individual with profound form vision deficits can still use visual texture and color to classify scenes—and that this intact ability is reflected in differential activation of the PPA with colored versions of scenes. ■

INTRODUCTION

Research on high-level visual perception has focused largely on object recognition. Yet the visual system must be able to recognize entire scenes, not only individual objects. The boundary, however, between what is a scene and what is an object can sometimes depend more on context than on content. For example, a desk may be pictured as one of several objects within an office scene, while at the same time the surface of that desk might be viewed as a scene in itself, one composed of various items on its surface—pencils, coffee cups, paper, books, keyboard, mouse, and monitor (Henderson & Hollingworth, 1999). For this reason, defining a "scene" can be a difficult task. Typically, a scene is defined as visual information about the immediate environment, often including a combination of both background elements and one or more discrete objects (Henderson & Hollingworth, 1999).

Biederman (1972) was one of the first to ask whether scene perception depends on the processing of individual objects or whether more global units or schemata are used. It certainly used to be commonly believed that scene perception is simply an extension of object processing. According to this bottom-up account, scene perception unfolds in the following manner: Basic features are extracted from the retinal array, individual objects are constructed from those features, and finally the objects are combined to deliver our perception of the overall scene. In other words, scene perception makes use of the same basic mechanisms used in object perception.

In contrast to this bottom-up view, other accounts of scene processing have argued that more global aspects of a scene are processed prior to or simultaneously with the identification of the individual objects composing that scene. Increasingly, evidence seems to support this view, namely, that scene perception is an independent process rather than an extension of object recognition. For instance, if scene perception depended on combining information from each of the individual objects within the array, then one might expect reaction times (RTs) for scene identification to be substantially slower than RTs for the identification of individual objects. (This does not exclude the possibility that visual information from several nonidentified objects in a scene might summate in some fashion and give rise to scene identification in the absence of object identification.) Moreover, one might also expect scene processing to occur in the direction of fine-to-coarse spatial scales, where fine details would be initially extracted to identify

[1]The University of Western Ontario, [2]Robarts Research Institute, [3]University of Durham

objects before the coarse scene information could be used to provide a global impression of the scene. But neither of these appears to be the case. As it turns out, scenes can be identified quite rapidly—as rapidly as the individual objects constituting the scene (Gegenfurtner & Rieger, 2000). Even more compelling is the fact that scene perception occurs quite effectively, and sometimes even preferentially, in the coarse-to-fine direction of processing (Oliva & Schyns, 1997; Schyns & Oliva, 1994, 1997). Indeed, scenes can be identified solely on the basis of coarse (low spatial frequency) information from extremely brief (50 msec) presentations, in which individual objects cannot be differentiated (Schyns & Oliva, 1994).

The idea that scenes can be identified on the basis of global properties alone has also received support from computational models that permit accurate scene classification without the prior identification of individual objects (Oliva & Torralba, 2001; Vailaya, Jain, & Jiang Zhang, 1998). Models that allow fast recognition of scene gist, without the identification of component objects, make use of properties such as the spatial complexity of the scene, visual texture, and color (Oliva & Schyns, 2000; Oliva & Torralba, 2001). Natural landscapes, for example, tend to have zones with undulating contours and characteristic textures and/or colors, whereas manmade scenes tend to have straight horizontal and vertical lines without such zones.

Behavioral studies have shown that color can be a particularly useful cue in the identification of natural landscapes. For example, Oliva and Schyns (2000) showed that color facilitates naming latency when scenes such as forests, deserts, and oceans are presented, but has no effect on naming latency for nonnatural scenes, such as cities or rooms. They also showed that naming latency increases for natural scenes when they are presented with inappropriate coloring or when all chromatic information is removed, whereas this is not the case for nonnatural scenes. Oliva and Schyns argued that color cues do not assist in the recognition of nonnatural scenes for two reasons: First, color patterns are less consistent in a nonnatural scene since the components of rooms and cities vary widely in coloration, and second, the color information that is available is found for the most part at a fine spatial scale. For instance, the color in a city scene might consist of the individual colors of the clothing of pedestrians or of the merchandise in a store window, or the light and signs along a street. But this fine-scale spatial information may not be initially accessible if fast recognition judgments are based primarily on the coarse-scale scene information. Natural scenes, by contrast, are composed of large regions of nearly uniform color that can be used as a cue for scene gist when only coarse-scale information is available. In fact, there is evidence that, in contrast to the discrimination of even simple geometric shapes, rapid categorization of scene gist requires little or no focal attention (Braun, 2003; Li, VanRullen, Koch, & Perona, 2002). In sum, behavioral testing of models of scene perception has shown that the scene gist can be extracted from coarse-scale information without the prior identification of particular objects in that scene.

Neuroimaging has provided further support that separate neural mechanisms are involved in object and scene perception. Functional magnetic resonance imaging (fMRI) has shown that the lateral occipital complex (LOC) is selectively activated during object processing (e.g., Grill-Spector, Kourtzi, & Kanwisher, 2001; Malach et al., 1995), whereas the parahippocampal place area (PPA) is selectively activated during scene perception (Epstein & Kanwisher, 1998). In fact, the PPA responds strongly to scenes but only weakly to objects and it does not respond to faces (Epstein & Kanwisher, 1998). Moreover, the PPA shows equal activation for images of empty rooms (depicting bare spatial layout) and furnished rooms (Epstein & Kanwisher, 1998; Epstein, Harris, Stanley, & Kanwisher, 1999). Its response to arrays of objects, such as furniture, without spatial context, is less than half of that for empty or furnished rooms. In addition, PPA activation is far stronger for intact rooms and fractured rooms (in which the walls and floors were separated by spaces but maintain the same spatial relation to one another) than for rearranged room elements that did not define a coherent space. In short, it appears that the PPA response is not due to the individual components of a scene's layout (the walls or the floor) but rather to the spatial layout as a whole.

Patients with damage to the PPA show impaired discrimination of novel scenes but not novel objects (Epstein, DeYoe, Press, Rosen, & Kanwisher, 2001). Patients with topographical agnosia, who show deficits in their ability to navigate in the environment, also have damage in this region, either in the PPA itself or in a region of the lingual gyrus immediately posterior to the PPA (Habib & Sirigu, 1987). Recent functional imaging has demonstrated that this lingual ''landmark'' area or LLA shows greater signal change for buildings than for faces or objects (Aguirre, Zarahn, & D'Esposito, 1998). In fact, differential damage of the PPA and LLA may explain the variability in the nature of the deficits in patients with topographical agnosia (see Aguirre & D'Esposito, 1999 for a review). Epstein et al. (1999) have suggested that the LLA may be involved in place recognition while the PPA may be more involved in memory encoding of place information—or alternatively, that both areas are involved, but in different ways, in the perceptual analysis of places, while memory for places is stored elsewhere. But in any case, the neuropsychological evidence is clear: Patients with damage to the PPA have difficulty identifying scenes (and in some cases, navigating through familiar environments) even though they have no deficits in recognizing objects.

In the present experiment, we examined an individual with quite the opposite pattern of deficits and spared

abilities. We showed that D.F., an individual with profound visual form agnosia, who cannot identify objects on the basis of their shape, can use her spared color and texture perception to classify scenes reasonably accurately. Using fMRI, we showed that D.F. had higher activation in the PPA for scenes presented in their normal color than for two-tone versions. Control observers showed the same level of activation in the PPA for all scene versions. These data provide further evidence that objects and scenes are processed by mechanisms that are functionally and anatomically independent—and that the identification of scene gist can take place on the basis of color and texture cues, without identification of object form.

RESULTS

Behavioral Studies of Scene Classification

D.F. is woman of 47 years, who suffered irreversible brain damage as a result of accidental carbon monoxide poisoning at age 34. D.F. shows relatively normal static perimetry in the central visual field up to $30°$ eccentricity with evidence of field loss only in the right lower quadrant. She has profound visual form agnosia (a deficit in object recognition based on form), which has been detailed elsewhere (Milner et al., 1991). D.F. has great difficulty perceiving the shape, size, and orientation of objects. Moreover, she is unable to copy line drawings of objects, although she can draw objects reasonably well on the basis of long-term memory (Servos, Goodale, & Humphrey, 1993). D.F. can discriminate among hues and name colors appropriately (Milner & Heywood, 1989). Consequently, D.F. can often recognize ''real'' objects, particularly natural objects such as fruit and vegetables, on the basis of surface information such as color and visual texture (Humphrey, Goodale, Jakobson, & Servos, 1994). For instance, she can correctly identify ''real'' natural objects such as a green pepper or an apple with 100% accuracy, but when presented with a line drawing of these objects, her accuracy drops below 10%. Similarly, she can recognize nearly 70% of ''real'' colored manufactured objects such as a blue running shoe or a yellow plastic dustpan but fewer than 10% when presented as a line drawing. Likewise, ''real'' achromatic manufactured metallic objects such as a fork or a camera were recognized with nearly 60% accuracy but line drawings of these objects resulted in approximately 10% accuracy. Naming latencies were significantly faster for real natural objects than latencies for colored manufactured and achromatic metallic objects which did not significantly differ from each other.

MRI carried out on D.F. one year after the accident revealed a distributed pattern of brain damage consistent with anoxia, but the damage was most evident in the lateral occipital cortex and the medial occipito-parietal region (Milner et al., 1991). More recent MRI scans show three distinct lesions, one in the lateral occipital cortex of each hemisphere and one in the left hemisphere near the top of the parieto-occipital sulcus with the primary visual cortex and the fusiform gyrus spared (see Culham, 2004; James, Culham, Humphrey, Milner, & Goodale, 2003). Figure 1 shows D.F.'s primary lesions in a structural MRI. The regions outside the three primary lesions remain functional despite widening of the sulci. For example, both D.F. and controls showed robust activation in the anterior intraparietal sulcus (AIP) during object grasping (James et al., 2003). The location in stereotaxic space (Talairach & Tournoux, 1988) of D.F.'s bilateral lateral occipital lesions overlap almost completely with the lateral occipital cortex (area LO) as defined by fMRI activation in normal observers viewing images of objects. In other words, D.F.'s lesions are localized in the very regions of the occipito-temporal cortex that have been implicated in the visual processing of objects in the normal brain.

D.F. and 19 control participants were shown a series of 600 pictures of common scenes in which images were presented in three categories of color-diagnostic (natural) scenes (beach, desert, and forest) and three categories of color-nondiagnostic (nonnatural) scenes (city, market, and room). The scene images were presented in five different formats: (1) normally colored, (2) color-inverted (i.e., each color was replaced by another color on the opposite side of color space), (3) grayscale, (4) black-and-white (i.e., two-tone), and (5) spatial rotation, in which the normally colored scene was rotated by $180°$. Examples of these different images are shown in Figure 2. D.F. and the control participants were asked to classify the scenes as quickly and as accurately as possible.

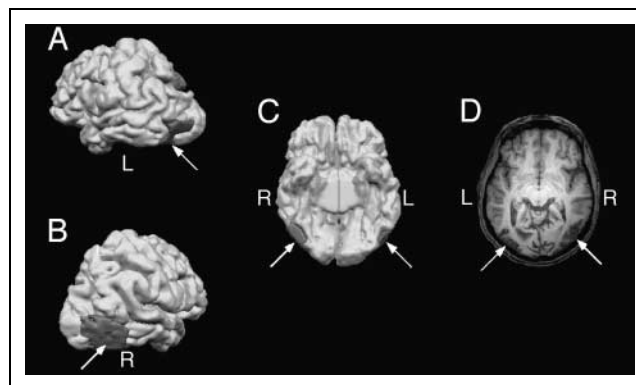A two-way within-subjects analysis of variance (AN-OVA) was carried out on control participants' vocal RTs

Figure 1. Ventral stream lesions in D.F. D.F.'s brain has been rendered at the pial surface (outer gray matter boundary) and is shown in A through C. Lesions were traced on slices that indicated tissue damage and rendered on the pial surface in dark gray. Arrows indicate the bilateral LO lesions. (A) Left hemisphere. (B) Right hemisphere. (C) A ventral view of the underside of the brain. (D) A transverse slice through the lesions ($z = -8$).

Figure 2. Examples of images depicted in the five different formats for each of the six scene categories.
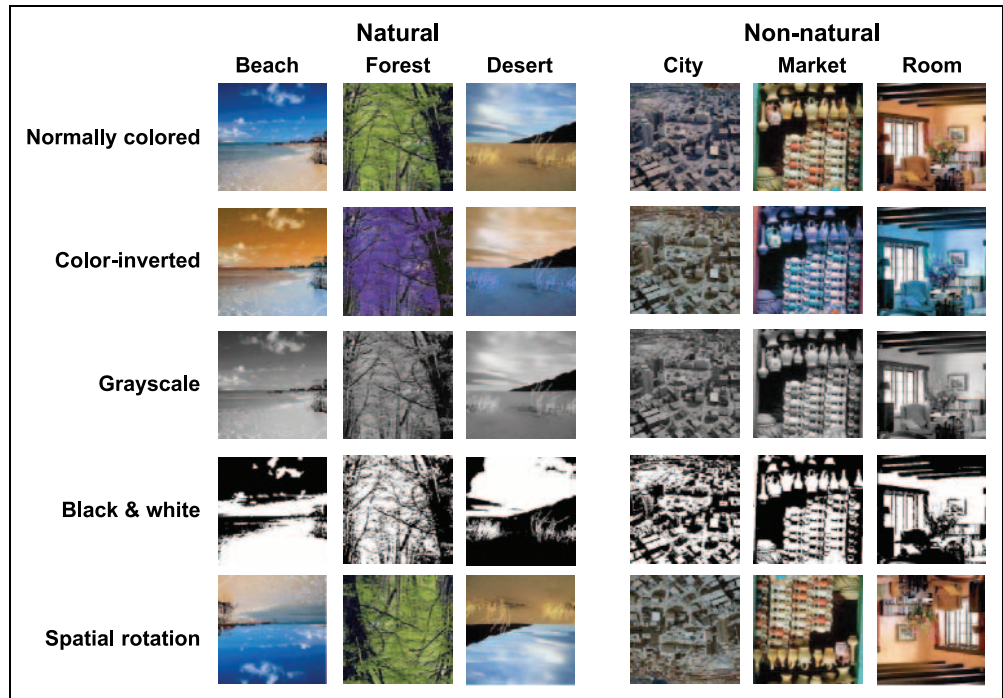
for scene category (natural vs. nonnatural) and scene format (normally colored, color-inverted, grayscale, black-and-white, and spatial rotation). This revealed a significant main effect of scene category indicating that, when averaged across scene formats, nonnatural scenes (such as cities or rooms) were recognized faster than natural scenes (such as beaches and forests) [$F(1,18) = 35.60$, $p < .01$]. The ANOVA also revealed a significant

Figure 6. Functional activation for scene images is shown in a coronal and transverse slice as well as on the pial surface in a ventral view of D.F.'s brain. PPA activation is shown by the area within the black box for each participant. Talairach coordinates in the left and right hemispheres are as follows: D.F.: −25, −50, −16 and 27, −52, −17; Control 1: −25, −50, −16 and 27, −61, −12; Control 2: −26, −49, −4 and 25, −51, −2; Control 3: −22, −55, −7 and 22, −54, −8.
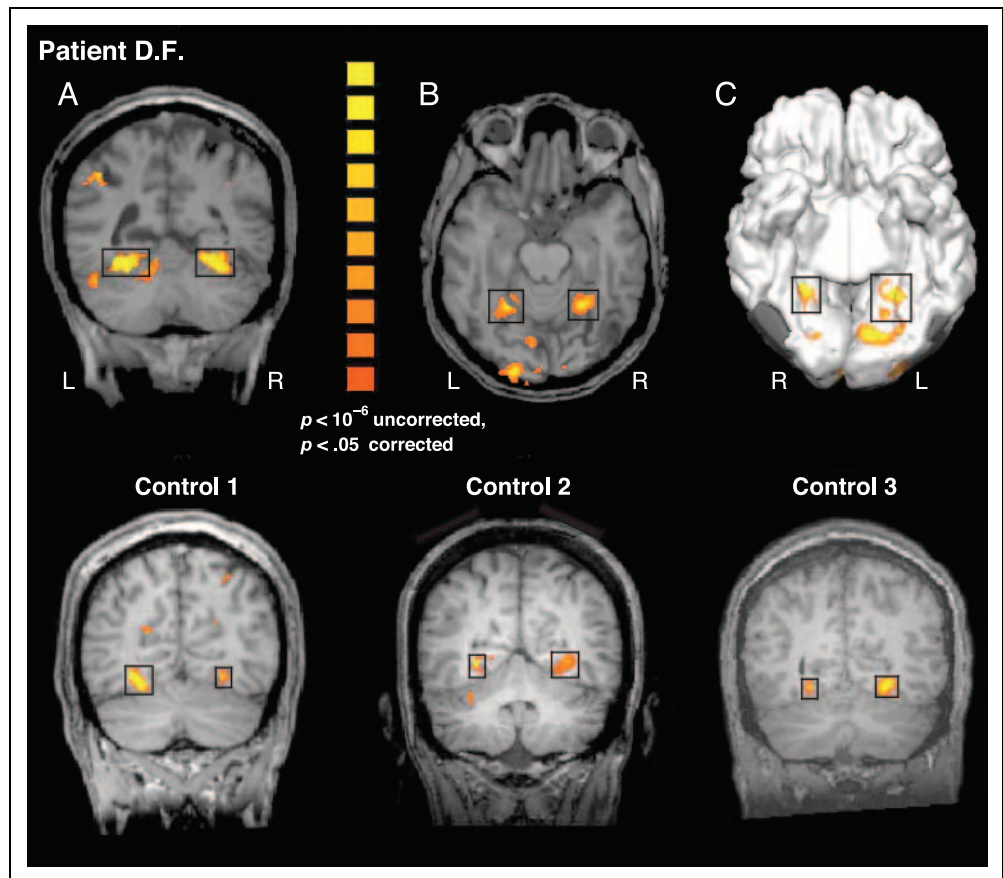
main effect of scene format indicating that, when averaged across scene categories, the format of the scene affected RT [$F(4,72) = 54.78, p < .01$]. In addition, there was a significant interaction of scene category with scene format.

The Scene Category × Scene Format interaction shown in Figure 3 is consistent with the data from Oliva and Schyns (2000): Chromatic information did influence scene categorization latency but not in the same way for natural and nonnatural scenes [$F(4,72) = 25.09, p = .01$]. This effect was revealed by a significant pairwise interaction between natural versus nonnatural and normal versus color-inverted (2 × 2 ANOVA) [$F(1,18) = 19.18, p < .01$]. Specifically, color inversion had a greater effect on categorization latency of natural than nonnatural scenes. Pairwise comparisons using Newman–Keuls post hoc tests showed that control observers had much slower RTs for color-inverted compared with normally colored natural scenes ($t = 9.31, p < .01$). This was not the case for nonnatural scenes ($t = 2.55, p > .05$). Additionally, there was a significant interaction (2 × 2 ANOVA) between scene category and color manipulation to black-and-white from the normally colored scene version [$F(1,18) = 44.43, p < .01$]. That is, the effect of making the image black-and-white had a much greater impact on natural scenes where color was diagnostic than on nonnatural scenes where color was not diagnostic—vocal RTs were much slower for natural than nonnatural black-and-white scenes ($t = 18.05, p < .01$). Tests of the pairwise interaction between the two scene

categories and manipulation to grayscale or spatial rotation were nonsignificant [$F(1,18) = 1.69$ and $1.54, p = .21$ and $p = .23$, respectively]. For natural scenes, pairwise Newman–Keuls post hoc tests showed vocal RTs were significantly slower for both grayscale and spatially rotated versions ($t = 3.58$ and $5.28, p < .05$ and $p < .01$, respectively) compared with those for normally colored scene images. Similarly, for nonnatural scenes, response latencies were significantly slower for grayscale and spatially rotated versions ($t = 4.93$ and $7.42, p < .01$, respectively) compared with normally colored scene versions.

Overall, abnormal color information had a greater influence on scene categorization for natural than for nonnatural scene categories. To put it another way, the addition of appropriate chromatic information increases the speed of categorization for natural scenes, whereas the addition of abnormal color information slows down this process. Taken together, these results illustrate that abnormally colored scenes interfere with the speed of categorization of natural scenes, where color is diagnostic, but not for nonnatural scenes, where color is not diagnostic.

We predicted that if scene processing and object processing are independent, then D.F. should be able to classify scene images. Further, because D.F. used color and texture information to recognize natural, manufactured, and achromatic metallic objects (Humphrey et al., 1994), we hypothesized that D.F. would also use color and texture information for categorization of both natural and nonnatural scenes. However, given that D.F. had greater accuracy for recognizing real (colored) natural than manufactured or achromatic metallic objects, we expected similar findings for natural and nonnatural scenes—that accuracy would be best and latencies shortest for normally colored natural compared to nonnatural scenes. Further, color manipulations such as converting a scene to grayscale or black-and-white should reduce performance for both scene categories as did line drawing versions of natural, manufactured, and achromatic metallic objects. Performance for grayscale versions should nevertheless be better than that for black-and-white versions because texture information is richer in the former versions. The prediction for color inversion is less clear—Color inversion should reduce performance for natural scene categories where color is predictive of scene category but it might also reduce performance for nonnatural scenes as did line drawing versions of manufactured objects. If D.F.'s spared color perception helps to define contours and boundaries comprising a scene whether color is inverted or not, color inversion of nonnatural scenes might give better performance than grayscale versions.

As predicted by a model of independent scene and object processing, D.F. was able to classify scenes relatively accurately although with much slower vocal RTs than the control participants (see Figure 4). For each
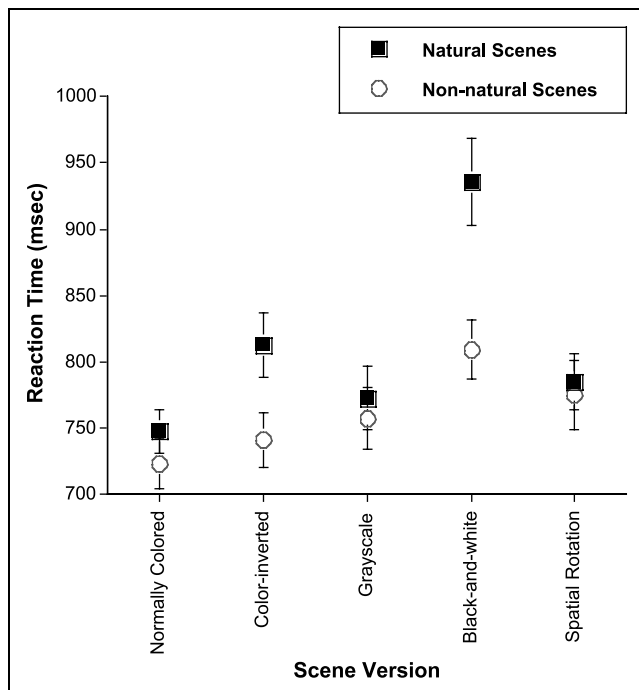

Figure 3. Mean vocal RTs (msec) for natural and nonnatural scenes in each of the five image formats averaged across all control participants. The error bars indicate standard error of the mean.
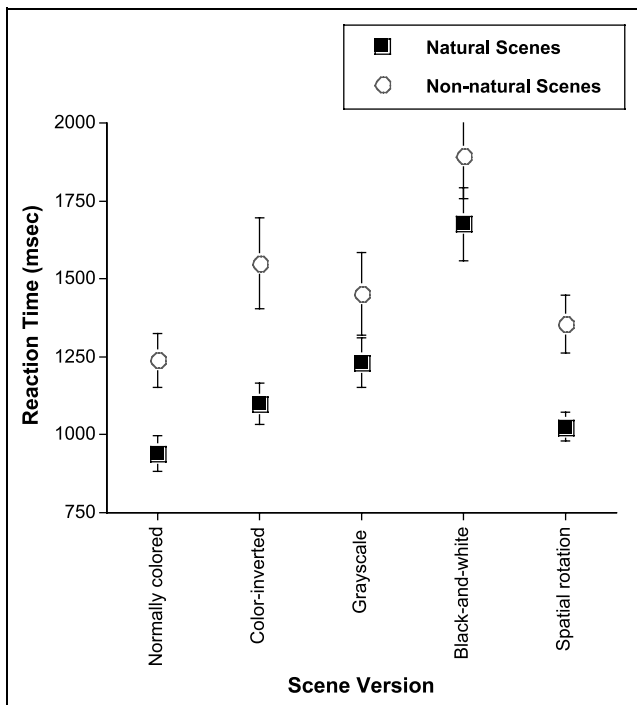
Figure 4. Mean vocal RTs (msec) for natural and nonnatural scenes in each of the five image formats for D.F. The error bars indicate standard error of the mean.

condition, missing values were replaced with the mean so that a repeated-measures ANOVA could be performed on the complete dataset. This revealed a significant main effect of scene category [$F(1,119) = 44.57, p < .01$]. RT for natural scenes was faster than it was for nonnatural scenes. The main effect of scene format was significant [$F(4,476) = 37.09, p < .01$]—manipulating the color of the scene significantly slowed D.F.'s RTs whether color was normally diagnostic or nondiagnostic. The interaction between color category and scene format was nonsignificant [$F(4,476) = 1.98, p = .10$]—unlike the control participants, D.F. showed the same pattern of RTs with different image formats for both natural and nonnatural scenes. In fact, her pattern of RTs for both kinds of scenes was similar to that observed with control participants viewing natural scenes in that latencies were longer for color-inverted compared to normally colored scene versions. These findings suggest that D.F. attempted to use color information to assist in scene categorization no matter whether it was diagnostic or not.

In general, control observers made few errors in categorizing scenes (mean = 1.5%), although more errors were made for color-inverted (2.5%) and black-and-white (5.0%) versions of the natural scenes. D.F.'s correct performance was well above chance (chance correct = 17%, chance incorrect = 83%) across all stimulus manipulations. She, however, made a large number of errors (mean = 17.3%), particularly for the black-and-white versions (27% for natural scenes and

53% for nonnatural scenes). Figure 5 shows error rates for controls and D.F. for each scene category and version. It is of note that D.F. made a great many errors of color misattribution—a kind of error rarely made by control observers. For example, D.F. routinely mistook color-inverted desert scenes (where sand becomes blue in the foreground; see Figure 2) for beach scenes (39%), and color-inverted beach scenes (where yellow becomes the predominant color; see Figure 2) for desert scenes (7.5%).

fMRI Investigations of Scene Categorization

During fMRI, D.F. viewed visual images of normally colored scenes, grayscale versions, black-and-white versions, faces, or a fixation stimulus by itself. As Figure 6 shows, like the three control participants we tested, D.F. showed significant activation in an area corresponding to the PPA when she viewed scenes as compared to when she viewed faces, a common contrast for defining activation in PPA (e.g., Epstein & Kanwisher, 1998). This was true for all versions of the scenes that were presented. The stereotaxic coordinates of the focus of activation in D.F.'s PPA were comparable to those seen in earlier studies of scene processing (Epstein & Kanwisher, 1998; Kanwisher, McDermott, & Chun, 1997).
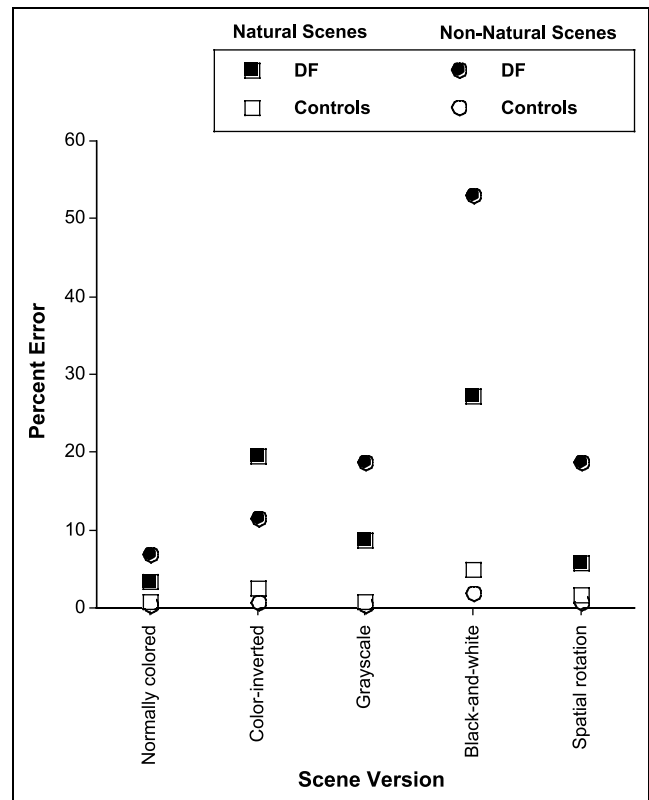

Figure 5. Categorization error rates for control participants and for D.F. for each scene type and image format. Chance is an error rate of 83.3%. The error bars indicate standard error of the mean for the control participants, some of which are too small to depict.

To compare differences in signal change between stimulus manipulations for each subject, we calculated the absolute difference in percent signal change for the contrasts between normally colored versus grayscale, normally colored versus black-and-white, grayscale versus black-and-white. We determined 95% confidence intervals around the mean of the three participants in the control group. The percent signal change for the normally colored versus black-and-white contrast for D.F.'s falls outside the 95% confidence interval of the mean of the controls. In other words, D.F. appears to show higher activation for normally colored versions of scenes than for black-and-white versions, unlike this group of control observers (see Figure 7).

DISCUSSION

The pattern of results from the behavioral testing of D.F., an individual with profound visual form agnosia, lends further support to the notion that scene perception can operate independently of object perception. In other words, the fact that she can classify scenes with considerable accuracy demonstrates that bottom-up object processing is not a necessary precursor to scene perception. Moreover, when D.F. was presented with scene images in the magnet, she showed robust activation in her PPA, an area that is thought to play a special role in scene analysis—despite the fact that she has large

Figure 7. Difference in percent signal change for the contrasts between normal versus grayscale, normal versus black-and-white, and grayscale versus black-and-white. Squares show the controls' mean difference condition contrast and error bars represent the limits of the 95% confidence interval around the mean. Diamonds show D.F.'s percent signal change difference for each condition contrast.

bilateral lesions in the lateral occipital area (area LO) of the LOC, an area implicated in object processing.

In the behavioral tests, D.F. was able to classify the majority of normally colored scene images correctly. It is possible that her performance may have been enhanced by her a priori knowledge of the six scene categories—but the same could be said for the control participants as well. In any case, she quite easily classified both natural and nonnatural scene categories and was even able to classify some of the black-and-white images correctly even though only limited visual texture information was available. In earlier studies, D.F. has also demonstrated that she is able to appreciate the visual texture and other surface features of objects such as specularities. When she is presented with real manmade objects, she often accurately describes their surface features and the material from which they are made—even though she cannot identify them or describe their shape (Humphrey et al., 1994). For example, when D.F. was shown a cheap flashlight sitting on a table, she said (correctly) that it was made of aluminum and red plastic, but guessed (incorrectly) that it was "some sort of kitchen utensil." As soon as it was placed in her hand, of course, she identified it as a flashlight. D.F. is also able to draw some of the textural features of objects when copying drawings, even though she does not render the overall shape of the object correctly and has no idea what it is she is copying (Milner & Goodale, 1995). It is likely that, if asked, D.F. would also be able to describe the colors and visual textures in scene images. It certainly seems likely that she is using such information to classify the scenes. The fact that natural scenes contain more predictable color, texture information, and spatial structure, than nonnatural scenes (Oliva & Torralba, 2001; Burton & Moorhead, 1987), may help to explain why she was able to classify natural scene versions faster than their comparable nonnatural scene versions.

Interestingly, D.F. demonstrated slower classification latencies for both natural and nonnatural scenes when they were inappropriately colored. This pattern of responding was quite different from that seen in our control observers for whom inappropriate scene coloring interfered more with classification latency of natural than nonnatural scenes. This suggests that, because D.F. has spared color perception, she makes use of color for scene categorization even when it is not always diagnostic for the scene. Thus, in nonnatural scenes, she may use color to parse the scene to some extent. Indeed, taken together, the results suggest that both global and local color may be important for D.F.; in other words, she makes use of color at a global scale in natural scenes and color at a local scale in nonnatural scenes. The importance of color at a coarse spatial scale for D.F.'s analysis of natural scenes is further supported by the pattern of specific errors that she made. Inappropriately colored desert and beach scenes were frequently confused for one another, probably because the color
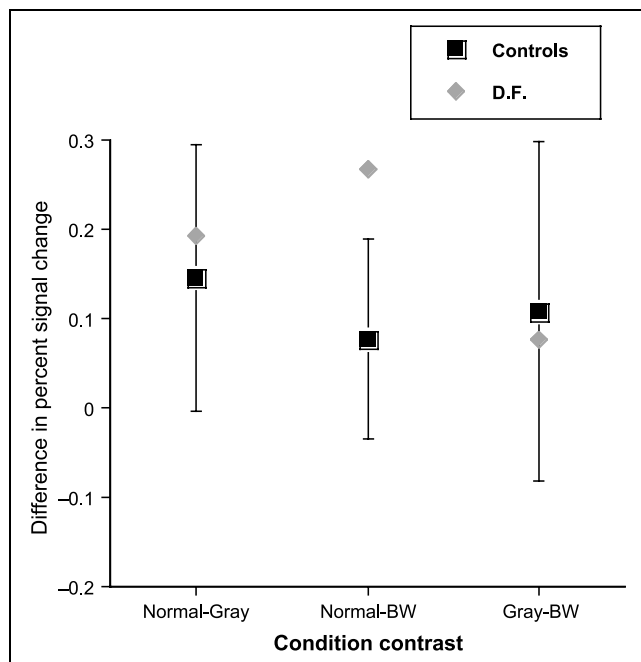
inversion caused the large predictable blue region in a beach scene to become yellow and conversely, the large yellow expanse of sand in a desert scene to become blue. D.F.'s poor performance with black-and-white scenes underscores the fact correct classification of these images depends heavily on the perception of form and contour information, a visual ability that is profoundly disturbed in D.F.

Functional imaging of visual areas in D.F.'s brain during passive viewing of scenes revealed that her PPA is still functional despite lesions encompassing area LO. In fact, D.F. appears to show more PPA activity for colored than for black-and-white scenes, a finding that is consistent with her performance in the scene classification task. This difference in activation between colored and black-and-white scenes was not present in our small population of normal control observers, a finding that is also consistent with their behavioral performance in the earlier scene classification task. The spatial aspects of form and color have been shown to be processed in separate, parallel neural pathways in the early stages of cortical visual processing (reviewed by Van Essen, Anderson, & Felleman, 1992). D.F.'s scene categorization performance is certainly less than normal, likely due to the disruption of some of the parallel pathways driving higher visual areas. Damage to one or more of these parallel pathways may impair higher-level visual functions but remaining parallel pathways may still drive these areas. Particularly because D.F. has retained normal color perception, inputs from color channels may drive these relatively larger changes in PPA response when viewing colored stimuli possibly as a result of a loss of other parallel channels coding more basic spatial structure. The fact that spatial form information plays a more important role in the classification of black-and-white scenes may explain why D.F. has much more trouble with these versions of the scene – and why her PPA tends to show relatively lower activation when she views black-and-white scene images. After all, she has large bilateral lesions of the object processing area LO and has great difficulty discriminating simple shapes or even telling vertical from horizontal lines (Milner & Goodale, 1995).

But the fact that the PPA shows at least some activation for the black-and-white scenes that D.F. had been unable to classify in the earlier behavioral task suggests that this region may be activated by the image of a scene even when it is not classified correctly. In fact, some investigators have argued that the PPA is not only involved in scene recognition per se, but may also play a more critical role in encoding local space into memory. Support for this argument comes from patients with damage to the PPA who show memory problems for spatial layout information and have difficulty navigating in an unfamiliar environment (Epstein, DeYoe, Press, Rosen, & Kanwisher, 2001). D.F.'s PPA activation for black-and-white scenes is consistent with this hypothesis; in other words,

she might have been trying unsuccessfully to encode the black-and-white scenes into memory. However, a conclusive test of this hypothesis would require a correlation of PPA activation with behavioral data during the scanning session as well as a comparison of PPA activation for black-and-white scenes with other nonscene objects which activate PPA but to a lesser degree than scenes (Epstein & Kanwisher, 1998). We can say, at least, that D.F.'s PPA activity correlates well with scene recognition in that fewer black-and-white scenes were recognized.

Taken together, our behavioral and imaging data lend further support to the notion that objects and scenes are processed quite independently. It is particularly interesting to note in this regard that PPA activation in D.F., unlike what happens in healthy observers, is modulated by the presence of color in the scene. This finding supports the idea that the PPA operates at a high, "categorical," level of processing, transcending the specific visual channels that provide it with relevant information. At a practical level, D.F.'s apparently intact PPA, and thus relatively spared place perception, may explain why she does not suffer from topographical agnosia. That is, her PPA seems able to cope with the limited amount of information it receives to put together a sufficiently complete cognitive map to enable her to get around her environment effectively.

This work also provides a compelling example of how studying visual processing in an individual with disordered vision can contribute to the identification of submodalities and independence in visual processing. In addition, combined behavioral and functional neuroimaging data from individuals with well-localized cortical lesions can shed light on the possible roles that different visual areas play in cognitive and perceptual processes.

METHODS

Scene Classification Task

Participants

D.F. and 19 male and female control participants, ranging from 19 to 53 years of age (mean age = 27 years; SD = 10.5 years), participated. All the control participants had normal or corrected-to-normal vision.

Stimuli

The stimuli consisted of three categories of natural or color-diagnostic scenes (beach, desert, and forest) and three categories of nonnatural or color-nondiagnostic scenes (city, market, and room). Twenty different scene images per category were taken from a CD photo image library. These 120 scene images were then each presented in five different formats for a total of 600 images: (1) normally colored: a true-to-life version of the scene; (2) color-inverted: Color hues were inverted by 180° in color space using Adobe Photoshop image software

(inverting the color hues produced the same color transformations as the code used by Oliva & Schyns, 2000, in CIE color space); (3) grayscale: The chromatic component of the image was discarded leaving a grayscale only image; (4) black-and-white: The contrast of the grayscale images was adjusted to 100%, which rendered a scene image where all regions were either black or white; and (5) spatial rotation: Normally colored scene images were spatially rotated by 180°. The size of all images was 256 × 256 pixels, subtending 4.3° × 4.3° when viewed from 150 cm. They were presented on an LCD panel display using a Macintosh G3 computer.

Apparatus and Procedure

Images were presented in 10 blocks consisting of 60 trials each. Within each, 12 scenes, two from each of the six categories (beach, desert, forest, city, market, or room), were shown in random order in each of five formats (normally colored, color-inverted, grayscale, black-and-white, and spatial rotation). A trial consisted of a 500-msec presentation of a red fixation dot (0.3°) followed by a 150-msec presentation of a blank screen followed in turn by the scene image. The scene remained visible until the participant responded. Trials were self-paced. Participants were instructed to name as quickly and as accurately as possible the category to which the scene belonged. Participants were given the category names before the experiment began. Verbal RTs were measured with a microphone that triggered the offset of the scene image. The experimenter coded the accuracy of each scene categorization without feedback to the participant. Participants were given a short practice run which showed only normally colored images. They were told before the first experimental run that they would be presented with images similar to the ones they had just seen as well as other images that could look a little "different." No feedback about performance was given throughout the experiment. For both control observers and D.F., only latencies for correct responses were included in the analyses.

Functional Neuroimaging Experiment

Participants

D.F. and three normal healthy control participants (control 1: male, age 34 years; control 2: female, age 46 years; and control 3: male, age 30 years) participated.

Stimuli

D.F. was presented with visual images of normally colored scenes, grayscale scenes, black-and-white scenes, faces, or a fixation stimulus alone. During the presentation of the scene images, free viewing was allowed. We especially wanted to examine activation for the condition in which D.F. had shown the worst performance: black-and-white versions of the scenes. We expected that D.F. might show differential activation for these scenes compared with those that she had correctly classified. Thus, the black-and-white images for the fMRI experiment were selected from the subset of black-and-white images that she had misclassified in earlier behavioral testing. The face stimuli consisted of color images of famous faces, eight men and eight women, seen from a frontal viewpoint on a black background. Faces subtended approximately the same retinal image size as the scene images (12°). (The face data are not presented here—see Steeves et al., 2003.) Each stimulus epoch lasted 16 sec, during which 16 different images were presented for 1 sec each. Each stimulus condition was repeated four times within each run (with a fixation period every fifth epoch) in pseudorandom order. The same 16 images were shown each time an epoch condition was repeated. Two runs were obtained on D.F. Because we had limited time to test this patient, we were unable to do an independent localizer scan to identify the areas that showed a greater response to scenes (places) than to faces or to include other conditions of interest (color-inverted or spatial rotation scenes). In order to maintain attention, both D.F. and the neurologically intact participants were asked to press a button when they perceived a "forest" scene. All participants, including D.F., completed the behavioral scene categorization experiment before the imaging session. They therefore would all have been exposed to the scene images before imaging and hence had a similar level of familiarity with the images.

For the control participants, we used an independent localizer scan to identify the scene area. The localizer run presented blocks of 16 sec of a fixation period alternating between blocks of either 16 face images or 16 scene images, presented for 1 sec each and repeated for four cycles. Each control participant was scanned during two or three experimental runs containing 16-sec epochs of the different scene conditions. Each run consisted of a 16-sec fixation epoch followed by blocks of 16 scene images (presented for 1 sec) in each of the five versions from the earlier scene classification task (normally colored, color-inverted, grayscale, spatial rotation, and black-and-white scenes) in three cycles. Sixteen different scenes were presented each time an epoch scene condition was repeated. Although D.F. saw fewer scene images and thus would have experienced greater fMR adaptation (Grill-Spector, & Malach, 2001), this cannot affect the interpretation of our data. That is, she showed PPA activation regardless of any adaptation and the degree of adaptation would have been equal across all conditions. Scene versions were presented in a quasi-counterbalanced order. Again, free viewing of scenes was allowed.

The stimuli were presented using a Macintosh G4 computer and a NEC VT540 projector at 800 × 600

resolution. Participants viewed the stimuli back projected on a screen through a mirror placed approximately 10 cm in front of the eyes.

Data Acquisition

Scans were conducted with a 4-T Siemens-Varian whole-body MRI system at the Robarts Research Institute using blood oxygenation level dependent (BOLD) imaging and a head coil for functional images. A series of sagittal T1-weighted scout images were acquired to select 16 contiguous, 5-mm-thick functional slices in a quasi-coronal orientation, sampling the occipital and posterior temporal cortex. Each functional volume was acquired using a navigator echo-corrected, slice-interleaved multi-shot (two shots) echo-planar imaging (EPI) pulse sequence with a 64 by 64 matrix size and a total volume acquisition time of 2 sec (TE = 15 msec, flip angle = 45°, FOV = 19.2 cm). Each imaging run consisted of 140 consecutive acquisitions of the selected brain volume. Within the same imaging session, high-resolution inversion-prepared 3-D T1-weighted anatomical images were acquired (64 slices, 256 × 256, 0.75 × 0.75 × 3 mm voxel size, TR = 9.8 msec, TE = 5.2 msec). In another session, participants were scanned using a cylindrical quadrature birdcage-style radio-frequency (RF) coil. Functional images were manually realigned to high-resolution anatomical images (1 × 1 × 1) that sampled the whole brain in order to obtain full-brain anatomical images to allow computation of stereotaxic coordinates (Talairach & Tournoux, 1988).

Image Analysis

Analyses were carried out using Brain Voyager 4.6 software (Brain Innovation, Maastricht, The Netherlands). Functional images underwent linear trend removal, spatial smoothing (gaussian filter, FWHM = 4 mm), and a correction for serial correlations. General linear model analyses were performed with separate predictors for each stimulus condition. Contrasts between predictors were used to identify regions of interest (e.g., + places, − faces). In control participants, a contrast between places and faces in the localizer scan was used to identify the PPA. Areas were defined as all contiguous activated voxels in the vicinity of the appropriate anatomical area that met a minimum threshold of $p < 10^{-6}$, uncorrected, $p < .05$ after a Bonferroni correction for multiple comparisons. Because no localizer scans in D.F. were available, we defined the PPA using the contrast between all three scene stimuli and the face stimuli. This contrast selected place-specific areas but did not bias subsequent analyses toward higher activation in any one of the scene conditions than the others. The map of the region of interest (PPA) from the independent localizer run was overlaid onto the experimental runs with different scene conditions in the con-

trols. Once regions of interest were defined, time courses were extracted from each run and the percent signal change was calculated using the fixation periods as a baseline. The activation levels for each condition were computed for each participant by averaging the percent signal change across all epochs within the condition.

REFERENCES

Aguirre, G. K., & D'Esposito, M. (1999). Topographical disorientation: A synthesis and taxonomy. *Brain, 122,* 1613–1628.

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to ''building'' stimuli: Evidence and implications. *Neuron, 21,* 373–383.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177,* 77–80.

Braun, J. (2003). Natural scenes upset the visual applecart. *Trends in Cognitive Sciences, 7,* 7–9.

Culham, J. C. (2004). Human brain imaging reveals a parietal area specialized for grasping. In N. Kanwisher & J. Duncan (Eds), *Functional neuroimaging of visual cognition: Attention and performance XX* (pp. 416–436). Oxford: Oxford University Press.

Epstein, R., DeYoe, E. A., Press, D. Z., Rosen, A. C., & Kanwisher, N. (2001). Neuropsychological evidence for a topographical learning mechanism in parahippocampal cortex. *Cognitive Neuropsychology, 18,* 481–508.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron, 23,* 115–125.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598–601.

Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology, 10,* 805–808.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41,* 1409–1422.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychologica, 107,* 293–321.

Habib, M., & Sirigu, A. (1987). Pure topographical disorientation: A definition and anatomical basis. *Cortex, 23,* 73–85.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Reviews in Psychology, 50,* 243–271.

Humphrey, G. K., Goodale, M. A., Jakobson, L. S., & Servos, P. (1994). The role of surface information in object recognition: Studies of a visual form agnosic and normal subjects. *Perception, 23,* 1457–1481.

James, T. W., Culham , J.C., Humphrey, G. K., Milner, A. D., & Goodale, M. A. (2003). Ventral occipital lesions impair object recognition but not object-directed grasping: A fMRI study. *Brain, 126,* 2463–2475.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences, U.S.A., 99,* 9596–9601.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., & Tootell, B. H. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Science, U.S.A., 92,* 8135–8139.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action.* New York: Oxford University Press.

Milner, A. D., & Heywood, C. A. (1989). A disorder of lightness discrimination in a case of visual form agnosia. *Cortex, 25,* 489–494.

Milner, A. D., Perrett, D. I., Johnston, R. S., Benson, P. J., Jordan, T. R., Heeley, D. W., Bettucci, D., Mortara, F., Mutani, R., Terassi, E., & Davidson, D. L. (1991). Perception and action in ''visual form agnosia''. *Brain, 114,* 405–428.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34,* 72–107.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41,* 176–210.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42,* 145–175.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5,* 195–200.

Schyns, P. G., & Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception, 26,* 1027–1038.

Servos, P., Goodale, M. A., & Humphrey, G. K. (1993). The drawing of objects by a visual form agnosic: Contribution of surface properties and memorial representations. *Neuropsychologia, 31,* 251–259.

Steeves, J. K., Culham, J. C., Duchaine, B. C., Humphrey, G. K., Schindler, I., Pratesi, C. C., Milner, A. D., Goodale, M. A. (2003). Functional activation of the fusiform face area and spared face categorization in a patient with dense prosopagnosia and visual form agnosia. Program No. 659.6. *2003 Abstract Viewer/Itinerary Planner.* Washington, DC: Society for Neuroscience, 2003. Online.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain.* New York: Thieme.

Vailaya, A., Jain, A., & Jiang Shang, H. (1998). On image classification: City images vs. landscapes. *Pattern Recognition, 31,* 1921–1935.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science, 255,* 419–423.