MATTERS
OF INTELLIGENCE

Conceptual Structures in Cognitive Neuroscience

Edited by

LUCIA M. VAINA
Harvard MIT Division of Health Sciences and Technology,
and Boston University

D. REIDEL PUBLISHING COMPANY
A MEMBER OF THE KLUWER ACADEMIC PUBLISHERS GROUP
DORDRECHT / BOSTON / LANCASTER / TOKYO
Chapter 3

Visual Texture for Recognition

Lucia Vaina

One of the most important tasks of the visual system in man is the recognition and identification of objects. An object within the field of view has place as well as form, color, size, texture, depth and motion. Many investigations have established the importance of shape for recognition. However, clinical and theoretical discussions of recognition of objects have tended to ignore the role of other visual properties in guiding recognition. This has been partly because shape information is usually more pertinent for manipulation purposes than any other object properties, but perhaps also because it is easier to think in terms of the geometry of spatial relations. Yet everyday visual experience tells one that other visual properties, such as texture, the pattern of an object’s material surface, can reliably guide recognition. For example, one can recognize a pineapple solely on the basis of the pattern of its skin, without needing to rely on additional information about its shape, size, or color. However, it is often harder to identify a lemon, an orange and even a cantelope only on the basis of their skin.


89

1 M. Vaina (ed.), *Matters of Intelligence*, 89-114.
(C) 1987 by D. Reidel Publishing Company.
pattern. How does this come about?

The image of a textured surface contains three dimensional information about the shape and distance of the surface relative to the viewer, and information about the texture itself such as its detailed structure and physical composition. It has been suggested that in early visual processing, visual texture is one of the important sources of information about the shape of the visible surfaces which is available in single images.

Early texture processes have as a goal the extraction of figure from ground, which is useful for scene segmentation and for describing object boundaries. It is plausible that in early vision, shape and texture need not be computed by independent processes, or described in different representations. However, although shape may be computed from smooth variation in texture, or texture gradient, it could also be computed from shading, contour, stereopsis and structure from motion. This suggests that impaired texture vision would not necessarily prevent shape description.

3.1 Does texture contribute in an important way to object recognition?

One way to think about the role of texture vision in higher level visual functions is as a two stage process: the first stage produces a unique description and the second uses it to identify objects.

The first question addressed by the present study is related to the nature of the representation in which texture information is described. Is this

---


4See Stevens, 1980.


representation semantically or perceptually based? A semantically based representation would entail that the classification of textures be dependent on one's knowledge of the object. A perceptually based representation, on the other hand, must be independent of previous experience knowledge and rely only on the information delivered by prior visual processing.

An approach to this question is to look at the nature of the errors produced by patients with localized brain lesions. It is well acknowledged that in individuals with standard cerebral dominance, the right hemisphere is more involved in processing higher level visual information of the kind which appears to be involved in perceptual categorizations—that is, a categorization of the visual stimuli according to some visual quality and which is independent of meaning or function. The left hemisphere, on the other hand, seems more involved in language and general conceptual categorizations which are associative in nature. Thus if visual texture were categorized perceptually, one would expect that patients with right hemisphere lesions would make more errors relative to perceptual categories, while patients with left hemisphere lesions will make more conceptual or semantic errors, in spite of processing a correct percept.

The second question addressed here refers to the way texture information might be organized by the representation. In the simplest case, no organization is imposed by the representation, and all elements in the description have the same status. Alternatively, the primitive elements could be organized into physically distinct modules consisting of adjacent elements of nearly the same size. For example, one module would emphasize the size and the local spatial arrangement of the hairs on an animal's fur. Another module would emphasize the spots and markings on the fur. Thus a modular organization would make possible the distinction between certain groupings from the others. It has been argued that in visual representations of shapes, these modules are organized from coarse to fine. If the same sort of organization held for texture vision, then different textures, such as that of a sheep and a cauliflower for example, often may look similar when they are compared at a coarse level. At finer levels which

---


emphasize the more local component elements of these textures, however, the individual hairs of the sheep fur will certainly help one differentiate it from a cauliflower. If this were true one would expect that recognition errors produced by patients with lesions to the brain which affect the performance of visual tasks could be differentiated by the degree of specificity of the level of processing. Thus textures which exhibit a more structured pattern, such as a pineapple, would convey enough specific information at a coarser level so that they should be more easily recognized than textures which have no further structure, such as leather or orange skin for example. Patients with lesions to the posterior right hemisphere would produce more perceptual errors than the left hemisphere patients on the more structured texture patterns such as textures which must be described hierarchically at different scales.

An experimental task was designed to address the following specific questions: (1) In processing visual texture for object recognition, is the first stage a perceptual categorization, which then is used to access the conceptual or associative representation of the object? (2) Is the texture representation hierarchically organized? Preliminary answers will be sought in the data obtained from patients with localized unilateral posterior lesions to the brain.

### 3.2 Subjects

Thirty-eight right-handed subjects with a mean age of 67.7 years participated in this study. Six were control subjects, matched for age, sex and handedness. (Table 3.1.) They were volunteers who came specifically to assist in research. None had any known history of cerebral disease or head injury. Thirty-two patients who underwent a single stroke and in whom a cerebral lesion was localized by CT scan and neurological consultation, constituted the subject population: seventeen subjects had right hemisphere lesions and fifteen had left hemisphere lesions. The lesions were located in the posterior temporal or parietal lobes or in the occipital lobes. All subjects had good language comprehension as determined by neuropsychological assessment. The left hemisphere patients with aphasia were anomic. One left hemisphere patient with aphasia was also graphic. None of the right or left hemisphere subjects were impaired in the recognition of visu-
ally presented shapes. They all were presented with a list of 15 silhouettes which they either could name correctly or at least indicate recognition in some other way (by paraphrase, or describing the function). None of the subjects presented impairment of stereo vision measured with the Julesz's stereograms or significant impairment of contrast sensitivity evaluated with the Vision Contrast Test System (Visitech, 1983).

### 3.3 Testing materials and methods

#### 3.3.1 Test stimuli

A series of familiar objects (plants, animals, man-made objects) was photographed under oblique morning lighting (45°) using Kodak Super XX Technicolor black and white film with a contrast index of .92. A section of the photographs was enlarged and the enlargements were cropped to exclude the identifying shape and size of the object (Figure 3.1). Twelve objects were photographed, and the corresponding cropped enlargements presented the texture of these objects. Eight of these textures exhibited various types and degrees of structure such as *tiled* (e.g. pineapple), *nested* (e.g. cauliflower), *woven* (e.g. sweater), *furry* (e.g. sheep, bear), *striped* (zebra, watermelon), etc. Two textures were *fine repetitive structures* (e.g. leather), and two had fine *grain structure* (e.g. elephant skin). (See Table 3.2.) The textures did not belong into mutually exclusive categories.

The perceptual categories were determined by asking five young normals with good vision and unfamiliar with the test and the responses, to judge a series of texture photographs and to group together those which they deemed to be perceptually similar to the targets. Additional three young normals were presented with silhouette drawings and were asked to group together those whose texture was similar. The test items were drawn from among those textures and silhouettes which were chosen in common by both groups of evaluators.

On separate 12 × 17 centimeter cards, silhouettes of objects were drawn in black ink. In each trial the subject was presented with a picture of a texture and a collection of five silhouettes drawings (Figure 3.2). The silhouettes were so chosen that in addition to the correct answer and an unrelated answer, there were two silhouettes whose texture was similar to
Table 3.1: Clinical Details of Patient Groups.

**A: Right Hemisphere**

<table>
<thead>
<tr>
<th>NR</th>
<th>Age</th>
<th>Sex</th>
<th>Hand</th>
<th>Site</th>
<th>Clinical Neglect</th>
<th>Aphasia</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60</td>
<td>F</td>
<td>R</td>
<td>P</td>
<td>LVFD</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>74</td>
<td>F</td>
<td>R</td>
<td>Occ</td>
<td>LVFD</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>57</td>
<td>F</td>
<td>R</td>
<td>P</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>62</td>
<td>M</td>
<td>R</td>
<td>Occ</td>
<td>LVFD</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>72</td>
<td>M</td>
<td>R</td>
<td>Occ</td>
<td>LVFD</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>62</td>
<td>F</td>
<td>R</td>
<td>P</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>65</td>
<td>M</td>
<td>R</td>
<td>TOcc</td>
<td>LVFD</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>62</td>
<td>F</td>
<td>R</td>
<td>Post</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>46</td>
<td>M</td>
<td>R</td>
<td>Par</td>
<td>LVFD</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>M</td>
<td>R</td>
<td>Occ</td>
<td>LVFD</td>
<td>No</td>
</tr>
<tr>
<td>11</td>
<td>65</td>
<td>M</td>
<td>R</td>
<td>TOcc</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>12</td>
<td>65</td>
<td>F</td>
<td>R</td>
<td>POcc</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>13</td>
<td>82</td>
<td>F</td>
<td>R</td>
<td>P</td>
<td>LVFD</td>
<td>No</td>
</tr>
<tr>
<td>14</td>
<td>77</td>
<td>M</td>
<td>R</td>
<td>P</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>15</td>
<td>56</td>
<td>M</td>
<td>R</td>
<td>Occ</td>
<td>LVFD</td>
<td>No</td>
</tr>
<tr>
<td>16</td>
<td>55</td>
<td>M</td>
<td>R</td>
<td>P</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>17</td>
<td>55</td>
<td>M</td>
<td>R</td>
<td>T</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

**B: Left Hemisphere**

<table>
<thead>
<tr>
<th>NR</th>
<th>Age</th>
<th>Sex</th>
<th>Hand</th>
<th>Site</th>
<th>Clinical Neglect</th>
<th>Aphasia</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>77</td>
<td>F</td>
<td>R</td>
<td>Occ</td>
<td>RVFD</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>59</td>
<td>M</td>
<td>R</td>
<td>TOcc</td>
<td>RVFD</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>62</td>
<td>M</td>
<td>R</td>
<td>Occ</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>61</td>
<td>M</td>
<td>R</td>
<td>P-T</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>57</td>
<td>M</td>
<td>R</td>
<td>P</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>M</td>
<td>R</td>
<td>T</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>67</td>
<td>F</td>
<td>R</td>
<td>P</td>
<td>VFD</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>66</td>
<td>M</td>
<td>R</td>
<td>Occ</td>
<td>VFD</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>60</td>
<td>M</td>
<td>R</td>
<td>Occ</td>
<td>VFD</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>62</td>
<td>M</td>
<td>R</td>
<td>MCA</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>11</td>
<td>55</td>
<td>M</td>
<td>R</td>
<td>POcc</td>
<td>mildRFVD</td>
<td>Yes</td>
</tr>
<tr>
<td>12</td>
<td>83</td>
<td>M</td>
<td>R</td>
<td>POcc</td>
<td>RFVD</td>
<td>No</td>
</tr>
<tr>
<td>13</td>
<td>57</td>
<td>M</td>
<td>R</td>
<td>POcc</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>14</td>
<td>63</td>
<td>M</td>
<td>R</td>
<td>POcc</td>
<td>RFVD</td>
<td>No</td>
</tr>
<tr>
<td>15</td>
<td>69</td>
<td>M</td>
<td>R</td>
<td>P</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Figure 3.1: Pineapple
that of the target, and one which was semantically similar to the target.

3.3.2 Testing methods

The textures and the silhouettes of objects were presented so that they were perceived in central vision. The first task involved perceptual matching. Using a multiple choice paradigm, the subjects were asked to identify the silhouette to which the target texture corresponded. For example, if the target was the texture of the pineapple (Figure 3.1), the correct answer entailed pointing to the silhouette of the pineapple in the multiple choice silhouettes presented in Figure 3.2.

The second task was semantic, and requested that subjects name the requested silhouette, or describe its function. The answers were scored as right or wrong. Failures were differentiated among perceptual, semantic or naming, and unrelated errors.

3.3.3 Results

Two types of scoring were used to analyse the results of the test, perceptual and semantic. Items were scored perceptually (perceptual matching tasks) based on the subject’s ability to correctly choose the object whose texture was given. Items were scored semantically (identification task) based on the subject’s ability to correctly name or describe by function the object he had chosen, whether or not it was correct perceptually.

Consider first the Perceptual Matching task. The errors were classified as (1) perceptual (e.g. matching the cauliflower texture with the silhouette of the sheep), (2) unrelated errors (e.g. matching the pineapple texture to the orange silhouette), and (3) no response. An error was categorized as perceptual if the subject chose a silhouette which depicted an object in the same perceptual category as the texture presented, although not correct. (See Table 3.2.) The errors of matching were scored as unrelated when the silhouette chosen was unrelated perceptually to the texture. In the identification task, three types of errors were possible, (1) the same class, which grouped those errors in which the response was semantically related to the object (e.g. orange for pineapple, or vegetable for cauliflower), (2) semantically unrelated (e.g. snake for pineapple), and (3) no response.
Figure 3.2: Silhouette Drawings.
Table 3.2: Test Stimuli: Structured and Unstructured Textures

<table>
<thead>
<tr>
<th>Structured Textures</th>
<th>Unstructured Textures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main Perceptual Feature</strong></td>
<td><strong>Main Perceptual Feature</strong></td>
</tr>
<tr>
<td>tile</td>
<td>pineapple</td>
</tr>
<tr>
<td>stripes</td>
<td>tire</td>
</tr>
<tr>
<td>furriness</td>
<td>bear</td>
</tr>
<tr>
<td>nestedness</td>
<td>cauliflower</td>
</tr>
<tr>
<td>woven</td>
<td>sweater</td>
</tr>
<tr>
<td>fissured pattern</td>
<td>tree bark</td>
</tr>
</tbody>
</table>
"No response" was not considered in the error analysis either for the perceptual matching task or the identification task. The task of matching texture to shapes presented no difficulty for the normal control subjects. (Only two errors were made; one was the choice of a lion instead of the bear, and the second, the leather was confused with the texture of an orange.) The identification task was also effortless for the controls. Naming of the objects was good (95%). The performance of the brain damaged subjects produced an interesting pattern. Table 3.3 presents the mean scores and the error analysis for the normals and the brain damaged subgroups.

A clear-cut difference is observed between the type of errors produced by the posterior left hemisphere patients and the posterior right hemisphere patients. Thus patients with left hemisphere damage made significantly less errors in choosing the object whose texture was given \( (t=3.36, \text{df}=30, p < 0.005) \); the mean number of perceptual errors in the right hemisphere group was \( X_{rh} = 5.1 \), while the number of perceptual errors in the left hemisphere group was only \( X_{lh} = 2.8 \). The mean number of correct answers in the left hemisphere group, \( X_{lh} = 7.3 \) and in the right hemisphere group was \( X_{rh} = 4.5 \). There was no significant difference between the right \( (X_{rh} = 1.4) \) and left hemisphere \( (X_{lh} = 1.5) \) brain damaged patients in the number of unrelated perceptual errors or no answer to the stimulus \( (X_{rh} = 1.1; X_{lh} = .5) \).

Since the right hemisphere patients produced more errors overall than the left hemisphere patients, the perceptual errors were further analyzed as a percentage of the total number of errors made (see Table 3.3). The result of the \( t \)-test showed that the percentage of perceptual errors computed over all the errors was significantly higher in the right hemisphere group than in the left hemisphere group \( (t = 2.073, \text{df}=30, p < 0.025) \). So far the data analysis suggests that subjects with lesions to the posterior right hemisphere performed significantly poorer on the perceptual texture task and made a significantly larger number of perceptual errors than the group of subjects with posterior left hemisphere lesions to the brain.

Looking again at the raw data (see Table 3.3) by subgroups, a puzzling fact emerged: no difference in performance on the perceptual matching task was obtained between the subjects with right and left occipital lesions. The puzzle is because at this level of the visual processing, in the occipital lobe, the incoming information is processed bilaterally, and hence one intact occipital lobe should allow the information to be processed accurately. Why
<table>
<thead>
<tr>
<th>Hemisphere or Group</th>
<th>Perceptual Mean Scores</th>
<th>Perceptual Error Analysis</th>
<th>Semantic Mean Scores</th>
<th>Semantic Error Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rights, N=17</td>
<td>42.2%</td>
<td>67.0%</td>
<td>Rights</td>
<td>42.0%</td>
</tr>
<tr>
<td>Lefts, N=15</td>
<td>60.5%</td>
<td>51.0%</td>
<td>Lefts</td>
<td>40.5%</td>
</tr>
<tr>
<td>Controls, N=6</td>
<td>97.2%</td>
<td>1.3%</td>
<td>Controls, N=6</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

**Error Analysis for Occipitals**

<table>
<thead>
<tr>
<th>Group</th>
<th>% Perceptual Errors All Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occipitals, N=10</td>
<td>73.8%</td>
</tr>
<tr>
<td>Others, N=22</td>
<td>54.5%</td>
</tr>
<tr>
<td>Left Occipitals, N=4</td>
<td>71.3%</td>
</tr>
<tr>
<td>Right Occipitals, N=6</td>
<td>75.4%</td>
</tr>
</tbody>
</table>

Table 3.3: Mean Scores and Error Analysis
wasn't there any difference between the right and left occipitals? All the occipital lobe lesions were produced by vascular accidents due to vertebral-basilar artery disease. The lesions were caused by thrombotic or embolic occlusions, and we hypothesize that the lesions may actually have resulted from ischemia in the occipital lobes.

Further statistical analysis indicates that the number of errors in the group of subjects with right occipito-temporal lesions was significantly higher than the number of errors produced by patients with the lesion situated in any other part of the posterior brain (OCCt > OTHER, t=2.81, df=30, p ≤ 0.0087).

The recognition of textures as instances in a perceptual category is perhaps a task which is carried out later in the visual processing, and it seems plausible that it be localized in the right hemisphere involving perhaps the most posterior part of the temporal lobe. These hypotheses need to be further investigated but it is conceivable that the specific tasks of shape and texture recognition may share some neural tissue.

On the identification task, there was no significant difference between the left and right hemisphere groups. (See Table 3.3, Semantic.) However, the analysis of the type of identification errors revealed a significant difference between right and left hemisphere subjects (LH > RH, t = 3.1754; df = 30, p < .005); the left hemisphere group producing more semantically related errors than the right hemisphere group. Specifically 39.2% of the identification (naming) errors produced by the left hemisphere subjects were semantically related to the object, compared to 18.3% related errors of the right hemisphere subjects.

### 3.4 Discussion

In this investigation, the performance of patients with left hemisphere and right hemisphere lesions has been compared on a task of visual recognition of objects through their textures. Two questions were asked: first, whether textures of objects are categorized perceptually or semantically; second, how is the information organized by the representation. The analysis of the results on the texture to object test showed that lesions to the left hemisphere could be associated with more conceptual (e.g., tiger instead of bear, or pineapple instead of cauliflower, saddle for horse), and naming
errors (the correct shape was indicated, but the subject named it wrongly; the subject pointed to cauliflower, but named it cabbage.) Individuals with right hemisphere lesions, on the other hand, produced more perceptual errors (sheep instead of cauliflower, and snake, fishnet or honeycomb instead of pineapple), naming correctly the shape they have incorrectly chosen. Thus this class of subjects could recognize shapes visually without difficulty, but they were impaired in texture recognition.

3.4.1 Categorical stages in texture processing

A pattern seems to emerge in texture recognition which differentiates errors into perceptual and semantical, and this suggests that visual textures may be processed in two stages: one which achieves a perceptual categorization (tiled structure), and the other which achieves a semantic categorization (the texture of a pineapple). Warrington and Taylor \(^{10}\) also proposed two serially organized categorical stages in their work on the visual recognition of objects. They demonstrated that the goal of the first stage is perceptual categorization, and it is carried out independently of the object's semantics. The second stage addresses the semantic categorization of the shape. In their study of three patients with visual agnosia, Ratcliff and Newcombe\(^{11}\) suggested that objects presented visually are identified on the basis of their visual characteristics alone, and that characteristics from other modalities, and especially the particular names of objects, are associated only subsequently. These empirical studies support Vaina's model of object recognition,\(^{12}\) which contains separate single modality descriptional representations for describing different aspects of an object, together with a functional representation useful for object manipulation and action comprehension. These various representations constitute separate processing modules within the model, occurring after the perceptual categorization, but before verbally mediated categorizations. Their goal in the visual modality is to provide descriptions which are not cluttered with the detailed mea-


\(^{12}\)See Vaina, 1983.
sures delivered by the early visual processes, but which capture the information in a form more immediately useful for recognition and manipulation.

Thus, it is possible that Warrington and Taylor's semantic stage consists in fact of two stages: one which occurs after the perceptual categorization (but preverbal), and its goal is to associate descriptions of the object's shape, texture, color and function in the visual modality, and the object's touch, smell, weight, taste and sound in other modalities. This is consistent with the recent results of Warrington and Shallice, which indicate that the semantic system is divided into material specific components. Thus, if an associative stage of processing which collects together the information from the single modal descriptions of the object such as its weight, color, shape and texture for example, followed the perceptual categorization stage, it would be conceivable that impairment of the perceptual categorizations, incomplete perceptual processing, or their intermodal connections could contribute to defective object recognition, and that the deficit may be selective depending upon the importance for recognition of those attributes which are not processed correctly.

The second stage of semantic processing is verbally mediated, and it contains the linguistic categorizations usually referred to as semantic memory. Impaired recognition, including impairment of object naming, may result from defective processing occurring in either of these preceding stages, or from deficits in their interconnections. It seems that the brain, after it takes the image apart and before it puts it back together into the whole object again, goes through different stages of processing which are separately susceptible to damage. Thus, first the image is analyzed in detail and its component characteristics are detected and symbolically described. This task is carried out in parallel equally in the most posterior parts of...
the right and left hemispheres.\textsuperscript{15}

Second, a \textit{perceptual categorization} occurs where more global characteristics such as shape, texture and color are described in special purpose modality-specific memory systems. Processes such as indexing into these memories and searching are especially important here. The clinical studies of brain injured patients suggest that the right hemisphere plays a dominant role in this stage of processing. Next, there is the associative stage where the perceptual characteristics are associated together in representations whose goal is recognition and identification of objects, and not merely their perceptual description. Here objects could be identified by their complex appearance and by their function or role in actions. Clinical literature seems to suggest that this stage is frequently carried out by the posterior left hemisphere. The \textit{verbal-semantic stage}, also carried out in the left hemisphere, involves associating names with objects, and this involves a \textit{verbally mediated categorization}.

\section*{3.5 Is this the whole story though?}

The above discussion suggests a serial organization of the visual processing:

- the analysis of the stimulus according to its different dimensions;
- the computation of structural percepts according to the object's salient properties such as shape and texture;
- the association of all the specific perceptual descriptions into a concept of the object, and finally,
- the verbal semantic stage where the object is associated with a name and is represented in a verbally mediated representation.

The present work seems to suggest that this order, however, need not be strictly respected.

Intriguing answers to the test stimuli were provided by a few subjects with occipital lobe lesions, and they deserve to be commented here. Two subjects (H.B. and D.V.) who had a basilar artery infarction (which, as we

\textsuperscript{15}See Warrington and Taylor, 1979; also L. Vaina, 1985.
have seen, suggests bilateral lesion) and a large lesion of the right occipital lobe, presented with the visual texture of elephant skin said immediately that it was a skin. One of them (H.B.), picked out the silhouette of an orange; asked if it could be something else, he said that it "could also be a lemon or a green pepper, because all these have skin". When H.B. was cued and asked whether the correct answer could be an animal, he said, "let's see, which animal has skin," and then decided that animals had fur, not skin, and thus it could not be an animal. Similarly, the subject I.M.—again with a basilar artery lesion—presented with the target of the elephant skin, said, "wrinkles, this must be thick skin". Then he looked at the choice of silhouettes and pointed to the elephant and said, "I cannot think of anything else."

It is possible that these subjects have associated as much as they have been able to process perceptually of the "elephant skin texture," not with shape but directly with a verbally mediated semantic property encoded in a linguistic name, "skin", which their by and large intact left hemisphere could handle correctly, and this name was used to index into the conceptual representation and search for objects which have "skin". The case N.S., when she looked at the cauliflower texture, said, "curly, curly", and then pointed to the sheep silhouette.

We could hypothesize that perhaps a semantic feature may also be associated with each specific type of texture, and that this feature is encoded in language, and it is given a name (e.g., wrinkled, skin, curly, etc.). Is perhaps the semantic level, which is more "stable" therefore more efficient for recognition? The perceptual descriptions which are carried out on the perceptual categorization level of texture processing are more accurate, and can deliver all the specific detail about the texture. But perhaps this isn't always necessary for object recognition and manipulation and it is possible that, in order to avoid combinatorial explosion of information, the passage from perception to cognition may involve a trade-off between accuracy and usefulness.
3.6 How is the information organized by the perceptual representation?

The way in which the information is organized by the representations dealing with the recognition of textured surfaces was the second question addressed in this investigation. We shall turn now to the discussion of the results of the texture to object test with the goal of unraveling the sort of organization that the representation might impose on the information in its descriptions. Two categories of test stimuli were used in the experiment. (See Table 3.2.) First, there were unstructured stimuli which allowed no further grouping or structuring. Unstructured textures, such as that of an orange skin or leather, are characterized by the simple distribution of some elementary tokens on the object’s surface, and no further grouping or structuring of these tokens could be achieved by subsequent processing. On these textures, both classes of patients with occipital lobe lesions performed poorly, and in general, the right hemisphere patients were more impaired than the left. The fact that no laterality effect was obtained with occipital lesions confirms the physiological data that early visual processing is carried out equally in both hemispheres.

The early processes are concerned with the detailed analysis of the image in terms of its primitive features. The recognition of unstructured textures could be considered to be more difficult because their differentiation requires discrimination among the individual elements which implies processing at a very fine spatial resolution. Processing at a coarser resolution is not informative because the texture elements do not present any specific groupings which may be detected at larger resolutions.

The second group of textures exhibited a more complex structure such as a pineapple skin or a cauliflower; and from looking at the errors produced by the right hemisphere patients in coupling the texture to the corresponding shape, a pattern emerged. The errors usually belonged to the same perceptual category with what would have been the correct answer (e.g., snake for pineapple). The errors produced by the left hemisphere patients were, as expected, more conceptual in nature. Thus, they tended to associate to the given texture the correct shape but in naming, they often retrieved a superordinate category (e.g., vegetable for cauliflower, or animal for bear) or a functional property (I eat it when I go out to the restaurant).
A possible explanation for these results could be that the computation of textures may occur at several different levels. The first level involves simple statistical measurements on the entire surface to determine what is there and how it is spread. For example, when one is looking at an orange or a canteloupe, simple density distributions and statistics of their primitive component elements indicate that there is a simple, uniform structure. Further processing of grouping and categorization could not extract additional information useful for recognition. However, if perceived at a very small, local scale, the two patterns could be differentiated because they have individual texture elements with different structures, although the specific distribution of the elements on the viewed surface is similar. Other textures, such as that of a pineapple or a cauliflower, are more complex. Here, grouping operations applied at different scales (spatial resolutions) may lead to the discrimination between various types of textures. Perceptually this implies that clustered texture elements could form regions which could be described visually at various scales. The scales used to describe the patterns occurring on a surface form a hierarchical description of the texture with the coarse, global statistical description on the top of the hierarchy and with descriptions using a finer scale at the bottom. This hierarchical description of the textures appears to be the organizing principle in a perceptual categorization which is independent of the object's specific form or of conceptual categorization of the object such as its function, semantic category or its name.

Thus, at the coarsest level, the texture of a pineapple, snake, fishnet and honeycomb are very similar;¹⁶ all four have a general surface structure characterized by the regularity and the repetitiveness of a similar hexagonal-like pattern. At a smaller scale, however, essential differences can be detected in the individual elements that form the pattern. These differences might be in the orientation or the exact shape of the component elements. Thus the fact that patients with lesions to the right posterior brain produced frequent errors which could be considered as belonging to the same perceptual category as the stimulus (sheep for cauliflower), appears to afford the explanation that these patients performed an incomplete processing of the stimulus. This processing was carried out at the coarser scales which were not sensitive enough to differentiate between textures belonging to the

¹⁶L. Vaina et al., "Perceptual categorizations of visual textures, in preparation."
same perceptual class. Individuals with intact posterior right hemisphere could process the perceptual information adequately, but they might have failed to use this information for indexing into the representation of the corresponding object presented at the conceptual level. However, the response was rarely all or none, usually some residual information was preserved, and, exactly as in the cases described by Warrington, this was of a more general sort.

The experimental results presented here suggest that texture information is represented at various scales organized from the more coarse to the more fine. This representation is constructed on perceptual criteria, independent of the meaning of the object. At a later stage of processing, however, these descriptions are put together with other visual and non-visual, intermodal, descriptions in the perceptually based component of the object's concept. To this, verbally mediated categorizations are added and thus the full concept of an object is achieved. The clinical data from patients with brain lesions suggested that perceptual categorizations are carried out specifically in the posterior right hemisphere, yet the association between these categorizations and the verbally mediated associations characterize left hemisphere functions. Anterior lesions were not included in this study. However, patients with frontal lesions on either side did not present any difficulty on the perceptual matching task.

3.7 Levels of processing in texture recognition: discussion in a computational framework

Experimental evidence was provided for the hypothesis that visual textures are described in a representation whose goal is the perceptual categorization of the texture types, and that this is independent of the objects to which they belong.

What is the nature of the descriptions that might be useful for recognition? Taking the pineapple skin as an example, different levels of organization may need to be made explicit. These are: (1) the specific texture of the individual component shapes on the pineapple's surface, (2) the outline of these shapes, and (3) the organization of these shapes (i.e. the repetitive
pattern they form on the surface).

Thus, two distinct levels of processing seem to occur in the description of textural patterns within a specific perceptual representation. First, the coarse, overall texture is described, without specific emphasis on the outline of the component shapes, or their organization on the surface. The second level computes the local contours of the repeated shapes, such as the hexagon on the pineapple or snake’s texture, or the long shape of the hair in the fur texture. The pattern of the repetition is also computed. Computationally, this entails a basic constraint on the design of a representation for visual textures,\(^ {17}\) that is, texture change contours should be made explicit in the image since they identify the likely location of the discontinuities in surface geometry or surface structure, and these may guide recognition. This constraint is shared by the processes whose goal is the computation of the shape, as well as texture. But what can we say about such a representation?

Psychophysics\(^ {18}\) tells us that the visual information is processed in parallel by a number of spatial-frequency tuned channels, which suggests that the visual system analyzes the image at multiple resolutions. Physiological experiments are consistent with the psychophysics, and they indicate that in the visual pathways spatial filters of different sizes operate at the same location. Receptive and dendritic field sizes of both retinal and cortical neurons increases monotonically with eccentricity, and this is consistent with the demonstration that the psychophysical channels depended on eccentricity.

For the computational analysis of vision, this suggests that the goal of the first category of processes must be to make explicit the intensity changes in the image, and these changes must be detected at different scales. Marr and Hildreth specifically developed a method for doing this.\(^ {19}\) They showed that the Laplacian of a Gaussian convolved image detection intensity changes at different scales in the image. They defined the operator

\[
\nabla^2 G = \left( \frac{\delta^2}{\delta x^2} + \frac{\delta^2}{\delta y^2} \right) G(x, y) = (1 - 4r^2/w^2)e^{-4r^2/w^2}
\]

\(^{17}\) See Marr, 1982.


where $\nabla^2 G$ is a two dimensional gaussian function, $w$ is the diameter of the positive central region of the operator, proportional to the Gaussian’s space constant, and $r^2 = x^2 + y^2$.

The operator is orientation independent, and thus it detects changes in intensity anywhere they occur within its resolution. The resolution of this operator, that is, the different scales at which the intensity changes are detected, depends on the value of $w$ and is spatially localized. The smaller operators detect elements with finer detail which often presents too much detail to afford successful recognition. Stronger physical characteristics of the viewed objects are detected by all sizes operators, and by and large, this information is more useful for recognition. Paralleling the neurophysiological results, Marr and Hildreth proposed that the descriptions delivered by applying individual operators, like the individual channel descriptions,\textsuperscript{20} are combined into a single description of intensity changes, in which contrast and width are made explicit. Thus not only is the information processed at different scales, but useful information is obtained by combining descriptions across scales.

What would be the role of these various sizes operators in the process of describing an object’s texture, such as the texture of a pineapple, for example? Larger $\nabla^2 G$ operators will show the overall organization of the surface of the pineapple, and perhaps the outline of the individual component shapes, but the textural characteristics of these component shapes would not be described. In other words, the size of the operator constrains the information made explicit by the process, and thus the information which is emphasized at any specific scale. The textured pattern of an object surface is essentially decomposed in zero-crossing segments which then provide the primitive texture elements. Similarities among these “texture” primitives are sought at the next level of processing, along some dimension such as orientation, contrast, or size.

Thus, Figure. 3.3 presents the textures of a sheep fur and a cauliflower. When the image of a sheep fur is convolved with the $G$ at various scales, one can see that the specificity of length and orientation of the individual component hairs are described more locally, and that larger operators will not emphasize this information. In fact when a larger operator is used ($w = 16$), only the rough groupings of large portions in the two textures are

\textsuperscript{20}See Campbell and Robson, 1978.
expressed. At that resolution, there is no essential difference between the cauliflower texture and the sheep texture. The differences occur at smaller resolutions where the individual hairs are made explicit in the sheep fur, yet they are not found in the cauliflower convolutions.

After the first "run" of the image with the Laplacian-of-a gaussian filters of different sizes, grouping processes must be used for the recognition of textures which present some structuring. The criteria for grouping may be suggested by the results of the statistical measurements on the image; the items with high frequency cause the texture analyser to group them along the prevalent dimension. What the exact dimensions are on which this grouping succeeds still needs to be elucidated. However, it could be hypothesized that because the spatial coincidence at all the scales of zero-crossings in the Laplacian of the intensity filtered with a gaussian mask reveals edges with specific orientations, and these edges have physical correspondence to the object, distinct from markings and shadows, it is possible that perhaps the further grouping processes would put together these edges in contours, and then in the larger pattern repeated on the viewed surface.

The recognition errors exhibited by patients with lesions to the right occipital lobe (reported in the previous section), interpreted in this framework lead to interesting testable hypotheses. Thus the high incidence of confusing the texture of a pineapple with snake skin, honeycomb or fishnet, suggests that the similarity of the component shapes of these patterns and the pattern of the repetitiveness of these shapes might play an important role in recognition. Indeed, in all three, the elementary shapes could be roughly approximated by an hexagon, which is then repeated regularly on the whole surface of the object.

A large operator of a low frequency applied to these images would produce similar results for all; differences among these textures would occur when $\nabla^2 G$ is used with a smaller width ($w$) which would process the images more locally. In other words, steep changes, are seen equally by all size operators, and these may determine the rough perceptual category of the texture. The hexagon-like component pattern and its repetitiveness is equally detected by all sizes masks. However, gradual changes are seen less well by larger masks, yet these gradual changes capture the specific differences between textures with similar gross structure.

Thus the interpretation of the experimental results and the result of the computational interpretation consistently suggest that in the recogni-
Figure 3.3: The images (sheep and cauliflower) convolved with a difference of Gaussian operator at three scales. The sign of the convolution is shown using white and black to indicate positive and negative regions respectively. The first pair (a1, a2) is with a $32^2$ operator ($w = 16$ pixels); (b1, b2) is with $20^2$ operator ($w=8$ pixels); and (c1, c2) is with a $10^2$ operator ($w = 4$ pixels). The convolutions are carried out in a pipelined convolver, designed and implemented by N. Larson and K. Nishihara.
tion process range and resolution are not both required simultaneously. High resolution, which allows the discrimination of the fine details on the textured surface, requires a short range for the process. Coarse resolution, which would account roughly for the overall organization of the textured surface, involves large range operators.

3.8 What does the hardware tell us?

A recent study of Tootell, et. al.\textsuperscript{21} brings experimental support to this hypothesis at the "hardware" level. They found experimentally a spatial columnar organization related to spatial frequency. In their experiment, the location of the cytochrome oxydase (CO) rich structures and the pattern of the 14c-2-Deoxyglucose (2DG) uptake was examined relative to various visual parameters presented. When a high spatial frequency pattern was shown binocularly at all orientations, 2DG intake was maximum around the cytochrome oxidase rich spots and minimal on the spots. The converse 2DG pattern was seen when the animal was shown binocularly a low spatial frequency pattern at all orientations. Their results were consistent with Livingstone and Hubel's finding that neurons in the cytochrome oxydase rich structures (which they call "blobs") lack orientation specificity, while the interblobs regions have a high orientation specificity.\textsuperscript{22}

Thus, it is plausible to consider the CO spots (the Livingstone and Hubel's blobs) to correspond to the large scale processing at a low resolution, while the interspots (interblobs) correspond to the high resolution processing of the image at a smaller scale. In favor of this hypothesis is the fact that the receptive fields of blobs are much larger than the receptive fields of interblobs. Anatomical evidence then tells us that visual texture is computed by two qualitatively different types of hardware: one, whose main accomplishment is range, processes the shading, and the general, more statistical aspects of the texture; while the other, whose goal is resolution, processes finer aspects of the surface which are orientation dependent, such as the specific hairs of an animal fur, the details of the pineapple skin, the

surface of a peach, and so forth. It is conceivable then that the output of the blob system is aimed at determining the perceptual category, while the interblobs address the specific instance in a category. This computation seems to be carried out in later processing, in the V2 where the interstripes appear to process end stoppings or terminators.\(^{23}\)

This seems to suggest that the processing of object textures for recognition follows serially the same route through V1-V2-V4 and IT, similar to the route followed in processing of shapes and colors. Strong evidence has been provided\(^{24}\) for the fact that these visual modules form a hierarchy which processes visual characteristics relevant and useful for object recognition. Within each of these modules the various types of information may be processed in parallel, as if each level in the hierarchy would contribute to the perceptual descriptions of objects’ color, shape and texture.

Acknowledgments: I am grateful to the medical and neuropsychology service of the New England Rehabilitation Hospital for permission to investigate and report my findings on patients under their care. I wish to thank Diane Bainbridge for her assistance in assessing patients and to Keith Nishihara for his help with running the convolutions of textures. I am grateful to J. Allman, M. Alexander, E.K. Warrington, H. Barlow, J. Maunsell and T. Poggio for their comments on earlier versions of the manuscript. The texture to object test was discussed in detail with H. Goodglass, whom I thank for his generous collaboration. This work was supported in part by NSF grant GC-A-321529 and NIH US-PSH Grant NS-06209.

[Lucia Vaina, Ph.D., Sc.D. is a member of the Division of Health Sciences and Technology at Harvard/Massachusetts Institute of Technology, Cambridge, Mass. and a Professor in the College of Engineering, Boston University, Boston, Mass.]
