Direct and Indirect Cooperation between Temporal and Parietal Networks for Invariant Visual Recognition

Isabelle Otto, Philippe Grandguillaume, Latifa Boutkhil, and Yves Burnod
Université Paris 6

Emmanuel Guigon
Ecole Centrale de Paris

Abstract

■ A new type of biologically inspired multilayered network is proposed to model the properties of the primate visual system with respect to invariant visual recognition (IVR). This model is based on *10 major neurobiological and psychological constraints*. The first five constraints shape the architecture and properties of the network.

1. The network model has a Y-like double-branched multilayered architecture, with one input (the retina) and two parallel outputs, the "What" and the " Where," which model, respectively, the temporal pathway, specialized for "object" identification, and the parietal pathway specialized for "spatial" localization.
2. Four processing layers are sufficient to model the main functional steps of primate visual system that transform the retinal information into prototypes (object-centered reference frame) in the "What" branch and into an oculomotor command in the "Where" branch.
3. The distribution of receptive field sizes within and between the two functional pathways provides an appropriate tradeoff between discrimination and invariant recognition capabilities.
4. The two outputs are represented by a population coding: the ocular command is computed as a population vector in the "Where" branch and the prototypes are coded in a "semidistributed" way in the "What" branch. In the intermediate associative steps, processing units learn to associate prototypes (through feedback connections) to component features (through feedforward ones).
5. The basic processing units of the network do not model single cells but model the local neuronal circuits that combine different information flows organized in separate cortical layers.

Such a biologically constrained model shows shift-invariant and size-invariant capabilities that resemble those of humans (psychological constraints):

6. During the Learning session, a set of patterns (26 capital letters and 2 geometric figures) are presented to the network: a single presentation of each pattern in one position (at the center) and with one size is sufficient to learn the corresponding prototypes (internal representations).

These patterns are thus presented in widely varying new sizes and positions during the Recognition session:

7. The "What" branch of the network succeeds in immediate recognition for patterns presented in the central zone of the retina with the learned size.
8. The recognition by the "What" branch is resistant to changes in size within a limited range of variation related to the distribution of receptive field (RF) sizes in the successive processing steps of this pathway.
9. Even when ocular movements are not allowed, the recognition capabilities of the "What" branch are unaffected by changing positions around the learned one. This significant shift-invariance of the "What" branch is also related to the distribution of RF sizes.
10. When varying both sizes and locations, the "What" and the "Where" branches cooperate for recognition: the location coding in the "Where" branch can command, under the control of the "What" branch, an ocular movement efficient to reset peripheral patterns toward the central zone of the retina until successful recognition.

This model results in predictions about anatomical connections and physiological interactions between temporal and parietal cortices. ■

INTRODUCTION

A particular object can correspond to an infinite number of possible retinal images. A bottle, for example, can be seen at different distances, in different positions, and with different shapes and colors. However, we can always recognize it. This is what we call invariant visual recognition (IVR).

This property of invariant visual recognition is of very great interest for both neuroscientists and engineers.

From a neurobiological point of view, IVR raises the question of what kinds of processes operate on the retinal image to allow access to previously stored "internal representation(s)" necessary for recognition. From a traditional engineering point of view, it would be extremely useful to design computers and robots that would be able to recognize objects in naturally varying perspectives and contexts. For this purpose, work on pattern recognition (particularly alphanumeric patterns) has focused on different methods:

1. Template matching: the simplest way to recognize patterns would be to match each incoming pattern against the set of stored templates. Not only does this method need a previous "cleaning up" of the image to normalize the patterns in size, angle, etc. but also it is not satisfactory for generalization processes.

2. Feature analysis: it would be more efficient to discriminate patterns by dealing with the combination of local features that can distinguish one from another. But descriptions of patterns in terms of a set of mini-templates, such as a Pandemonium system adapted from Selfridge (1958), will fail on capturing overall structural relations. It will confuse patterns built up from common local features but globally organized in a different spatial arrangement (like a "T" and an "L").

3. A more flexible representation of patterns is provided in humans by language: a set of symbolic propositions can describe the components of a pattern, making explicit the structural arrangements of these primitives. This symbolic approach has often been attempted in artificial intelligence but the main problem lies in the necessity for humans to make an a priori cognitive description of what is pertinent to recognize a particular pattern and what is not.

More recently, artificial neural networks approaches have tried to get rid of these top-down inferences. It is the network itself that provides, by learning from examples, the adequate configuration of local primitives to discriminate each pattern from the others. Various global properties such as parallel data processing, noise suppression and associative recall can be used to improve performance. However, IVR is typically not an emergent property of these networks and thus particular models have been proposed to improve invariant visual capabilities: combination of AND/OR functions with increasing receptive field size (Fukushima, 1980; Fukushima & Miyake, 1982; Fukushima, Miyake, & Ito, 1983), graph matching with short-term plasticity (Bienenstock & Von der Malsburg, 1987), and networks with sigma–pi units allowing autocorrelations (Glünder, 1987).

However, these models do not tackle the problem of *invariant recognition such as that faced by the human visual system* and they do not mimic its powerful performances: for patterns presented with widely varying sizes, textures, colors, and orientations (for familiar objects), visual identification can be achieved on the basis

of only approximately 100 msec of processing (Thorpe, 1989; Thorpe & Imbert, 1989). For learning and identification of patterns, it is important to take into account the sensorimotor loop capturing the object in the center of gaze: when a composite pattern is presented to a subject in varying positions in the visual field and ocular movements are not allowed, it is recognized only in the restricted region where it has been learned (Lévy-Schoen, 1977; Walker-Smith, Gale, & Findlay, 1977; O'Regan & Lévy-Schoen, 1983; O'Regan and Nazir, 1990). To investigate the invariant capacities of the visual system, it is thus important to associate in the same network both direct sensory perception and adaptive sensorimotor control of gaze.

In this article, we propose a neural network that models specific biological properties which could be responsible for the IVR capabilities of the human brain. This model is not based on a unique factor but takes into account the most striking features of the visual system of the primates that we have summed up by *10 major psychological and neurobiological statements.* The first five constraints operate on the functional architecture of the network and the other five concern its expected behavior and performances.

We will first expose these 10 statements and then detail their biological validity and explain how they shape the neural network model.

Functional Architecture

Segregation between Two Visual Pathways

The multiple visual areas are organized into two separate cortical visual pathways: a ventral stream specialized for "object" vision and a dorsal stream for "spatial" vision. The divergence between these two systems appears to begin after striate cortex.

We thus propose a network whose global architecture is Y-like shaped with a "What" branch performing object recognition and a "Where" branch encoding the location of the pattern in the peripheral field. We have tested and compared the performances of each branch separately on one hand and of the two cooperating branches on the other hand.

Number of Steps and Their Respective Function

The visual system transforms the pattern of excitation that reaches the eyes from a retinotopic coordinate system to a coordinate system centered on the object itself. This final representation is achieved by a forward progression from striate cortex to temporal lobe through a rather limited number of successive steps of information processing.

We have taken into account *the essential functional steps necessary to achieve this transformation,* with (1) first steps performing low-level processing of retinal information, (2) high-level terminal steps storing the

learned prototypes, and (3) intermediate steps learning the association between internal and external representations.

Central vs. Periphery Representation and Receptive Field Size

Receptive field (RF) size increases regularly in the successive steps of the two functional pathways. In addition, RF size increases as a linear function of eccentricity. These facts, coupled with the observation that the representation of the central vs. peripheral visual field is different in the two pathways (in the ventral stream areas, there is a larger representation of the central region whereas in the dorsal stream the peripheral region prevails) imply that RF size increases more rapidly in the dorsal stream than in the ventral one.

We relate this distribution of RF sizes in the step where learning takes place with the invariant capabilities of the system without ocular movements. We propose a solution for the necessary tradeoff between discrimination capabilities (in terms of fine grain construction of visual image) and invariant capabilities in both size and position.

Coding Properties of Neurons

Cortical neurons exhibit differential selective properties along the processing pathway: neurons are selective (1) to simple features like oriented edges, color, direction of movement of moving patterns in the first steps, (2) to complex patterns (or "prototypes") in the highest steps (and not to their component features), and (3) to both in intermediate associative levels.

We have implemented these three levels of coding. In the higher levels prototypes are coded by local populations of cells in a "semidistributed" way. In the intermediate associative steps, cells code a combination of both low-level characteristics (through feedforward connections) and high-level properties (through feedback ones).

Cortical Column-Like Processing Unit

Different types of cells coexist in the cortical tissue and form local circuits within the depth of the cortex with anatomofunctional characteristics organized in "columns."

In this model, the basic processing units are not single cells but model these basic circuits. We propose learning and activation rules that are consistent with the known properties of these neuronal circuits.

Expected Behavior and Performances of the Network

We expect from the model five behavioral performances resembling those of humans (and more generally of primates).

Immediate Learning

Humans and primates are capable of learning with only one trial.

A single presentation of each pattern in one position and one size only should be sufficient for the network to learn it.

Immediate Recognition of Centered Patterns

Humans can identify patterns centered on the foveal region on the basis of only 100 msec of processing.

We expect an immediate recognition in the temporal branch for centered patterns presented in the learned size (direct temporal recognition).

Moderate Size Invariant Properties

Without any active process modifying the size of the presented patterns, humans can recognize patterns in a given range of sizes, even if it has been previously presented once at one particular size.

The network should be resistant to changes in size of the presented patterns, at least for a limited range of sizes.

Moderate Shift-Invariant Properties without Ocular Movements

In a similar way, humans show shift-invariant recognition for patterns limited to a central region of the retina (again when presented centered and only one time in eye-fixed conditions).

We expect that, when ocular movements are not allowed, the recognition capabilities should be unaffected by changing positions around the learned one.

Cooperation between Recognition and Control of Gaze

Limits of recognition in eye-fixed conditions are overcome by the cooperation between the sensory and the oculomotor systems: recognition is achieved by the command of appropriate eye movements that can successively foveate on different objects (or different parts of composite objects). The information on object location information provided by the parietal lobe is an appropriate source for this oculomotor command.

When the pattern is not directly recognized by the "What" branch, the "Where" branch should be able to

command an eye movement that can reset the pattern in the center of the visual field where it can be recognized by the "What" branch.

We will now explain in detail how the five architectural and functional properties that are the most salient known features of the cortical visual system do constrain the network architecture and then show that the five behavioral expectations can be obtained using such a constrained model.

FUNCTIONAL ARCHITECTURE OF THE NETWORK

Modeling the Two Visual Pathways

The Two Visual Pathways in the Macaque Monkey

In the mammalian visual system, visual information is first processed in the retina and the lateral geniculate nucleus (LGN) by low-level operations (such as enhancement of spatiotemporal contrasts) before reaching the cerebral cortex.

Numerous neuroanatomical studies coupled with physiological recordings suggest that, among the 20 or so visual cortical areas devoted to vision (Van Essen, 1985), two major visual pathways originating from the striate cortex can be distinguished, dealing with two different kinds of visual information (Fig. 1A) (Ungerleider & Mishkin, 1982). One pathway, called the "Where" pathway, leads dorsally to the parietal cortex and is involved in motion detection, spatial location, and three-dimensional (3D) relationships. The other pathway, called the "What" pathway, leads dorsally to the inferotemporal cortex (IT) and deals with 3D form identification, surface properties, and color extraction (Zeki, 1973; Ungerleider & Mishkin, 1982; Mishkin, Ungerleider, & Macko, 1983; Van Essen & Maunsell, 1983; Shipp & Zeki, 1985; Desimone & Ungerleider, 1986; Maunsell & Newsome, 1987; Perret, Mistlin, & Chitty, 1987; De Yoe & Van Essen, 1988; Morel & Bullier, 1990).

Temporal lobe lesions can produce specific deficits related to object recognition such as an inability to recognize faces (Meadows, 1974a,b; Pearlman, Birch, & Meadows, 1979; Damasio, Damasio, & Van Hoesen, 1982; Joynt, Honch, Rubin, & Trudell, 1985). In contrast, lesions of the parietal cortex in Macaque monkeys lead to clear deficits on learning of tasks involving the recognition of relative spatial positions of objects in the visual field (Pohl, 1973). Furthermore, clinical observations in humans indicate that damage to the posterior parietal cortex can produce deficits including the inability to localize visual targets or to perceive movements, disturbances in the spatial distribution of attention, loss of spatial memories, and the inability to represent spatial relations in models or drawings, yet leave object recognition unimpaired (Ratcliff & Davies-Jones, 1972; Zihl, Von Cramon, & Mai, 1983).

The two kinds of deficits, produced by inferotemporal and parietal lesions, can be reproduced by disconnecting these two regions from the striate cortex (Mishkin, 1966; Iwai & Mishkin, 1969; Cowey & Gross, 1970; Mishkin, 1972; Ungerleider & Mishkin, 1982) demonstrating the importance of the *pathways,* i.e., of the successive cortical areas, linking the striate cortex to these two higher level processing regions. It has been suggested that the "What" pathway runs ventrally through areas V2, V4, TEO, and TE in the inferortemporal cortex while the "Where" pathway leads dorsally to area 7a or PG of the parietal cortex (Desimone, Fleming, & Gross, 1978, 1980; Ungerleider & Mishkin, 1982; Mishkin, Ungerleider, & Macko, 1983).

Although inferotemporal and parietal cortices receive largely separate prestriate afferents, they also exchange direct connections and both receive common inputs from extensive regions within the fundus of the superior temporal sulcus (STS) (Maunsell & Van Essen, 1983; Ungerleider & Desimone, 1986; Shipp & Zeki, 1985, 1989; De Yoe & Van Essen, 1985; Cavada & Goldman-Rakic, 1989a,b; Kuypers, Szwargbart, Mishkin & Rosvold, 1965; Seltzer & Pandya, 1980; Shiwa, 1987; Desimone, Fleming & Gross, 1978, 1980; Morel & Bullier, 1990). However, connections within and between the two streams are not of the same type (according to the laminar distribution of the labeled cells), the segregated prestriate connections being "feedforward" while the connections from STS areas are either "lateral" or "feedback" connections (Morel & Bullier, 1990).

The Network Architecture: A Y-Like Structure

A single multilayer network does not fit this branched architecture. We thus propose a Y-shaped network, with one input (the retina) and two outputs: the "What" branch models the ventral stream and the "Where" branch models the dorsal stream (Fig. 1B).

The retina (RETINA) is the unique input layer on which visual stimuli are presented while the model provides two different outputs: the "What" branch will signal for the recognized prototypes and the "Where" branch will determine target location in space (with a control on "eye" movements). These two branches are linked together by lateral connections that can control the interactions between the two processes to increase recognition capabilities.

Functional Steps

Successive Steps in the Two Pathways

Anatomical and physiological studies have provided important insights in the successive functional steps operating along the two pathways.

The inferotemporal cortex can in fact be divided in two anatomically and functionally distinct areas (Fig. 1A): (1) the anterior portion (Von Bonin and Bailey's area TE) involved in mnemonic functions that is hypothesized

B

"WHERE" PATHWAY   "WHAT" PATHWAY

pSA   X LOCATION   tSA   p PROTOTYPE

pAA   XY   tAA

SSA

PSA

RETINA

A

Parietal

IPS   PP   PO   LS

MST   MT   V2   U1

STS   V4   IOS

TEO

TEp

TEa

InferoTemporal

Figure 1. *The Y-like architecture of the multilayered network and the connectivity scheme. (A) Visual areas of the primate:* visual information is first processed by V1 and V2 and then by two major pathways: a ventral pathway toward the inferotemporal lobe rather involved in object recognition and the dorsal pathway toward the parietal lobe rather involved in localization and motion detection. The figure shows some important functional steps in each stream: V4 and IT subdivisions (TEO, TEp, TEa) in the ventral pathway, PO and PP (including 7a and LIP) in the dorsomedial pathway, and MT and MST in the dorsolateral pathway. *(B) The model:* The global architecture of the network is Y shaped with a "What" pathway performing pattern recognition and a "Where" pathway encoding the location of the pattern. Patterns enter the visual pathway through a 32×32 retina (binary coding). The two first "low-level" maps, common to the two streams, are called primary sensory area (PSA) and secondary sensory area (SSA). The two "high-level" maps are called temporal associative area (tAA) and temporal semantic area (tSA) in the "What" branch and parietal associative area (pAA) and parietal semantic area (pSA) in the "Where" branch. Each processing unit in these maps is represented by the smallest square (their inner divisions are not represented in this figure) and is labeled by three indices (X,Y,i) in PSA and SSA maps and by four indices (X,Y,i,p) in the associative and semantic areas. The figure suggests a possible 2D mapping of these indices on the surface of the cortical maps: the first two maps are retinotopic (their main axes are topographic: X,Y) with "local feature indices" (i). The topographic indices are maintained as main axes of the maps in the "Where" pathway although the prototype index "p" becomes the main axis in tSA (and thus with a loss of topography). Units in area PSA [$16(X) \times 16(Y) \times 4(i)$] performs an OR operation on the result of the filtering of a 7×7 RF by a 3×3 specific

orientation mask [i, with horizontal (H), vertical (V), and 2 obliques (D1) and (D2)]. Units in SSA [$8(X) \times 8(Y) \times 6(i)$] code for combinations of orientations (AND operation) extracted in PSA on a 5×5 RF by a 3×3 mask (OR operation), with six possible combinations (HH,VV, D1D1, D2D2, HV, and D1D2). Units in the two associative maps [$4(X) \times 4(Y) \times 6(i) \times 28$ (p)] depend both on the 3 indices of the feedforward inputs (X,Y, i) and on the "prototype" index "p" of the feedback inputs coming from semantic areas (28 prototypes are learned). "Feature" index (i) is the same in pAA as in SSA (combination of orientations), and within each index "i," a unit of area pAA performs a convolution of a 5×5 RF on SSA by a 3×3 mask (OR operation). Each group of cells with the same "prototype" index (p) in pAA is connected in a reciprocal mode with the cluster of units coding for the same index "p" in pSA. The same mapping is applied for connections from SSA to tAA but on smaller receptive fields size. The global number of units in tAA is the same as in pAA. Units in the semantic areas [$2(X) \times 2(Y) \times 6(i) \times 28(p)$] are connected in a reciprocal mode with units of similar "prototype" (p) and "feature" (i) indices in the associative area, with a convergence on the topographical indices (X,Y). RF greatly overlap in the "Where" pathway but not in the "What" pathway. The representation of prototypes in the temporal semantic area is thus "semidistributed" (among $2 \times 2 \times 6$ units). The position of the pattern in the visual field is coded in the parietal semantic area by a population of units (each one specialized for upper left, upper right, lower left, or lower right of the visual field). The two pathways are connected: a signal from tSA to pSA can influence (by gating or inhibition) the output of pSA which drive ocular saccades (see Discussion section). Only feedback connections from the semantic areas down to the associative areas are modified by learning.

not only to synthesize the analyzed attributes into a unique configuration but also to work as the storehouse for central representations of the objects (Miyashita, 1990) and (2) the posterior portion (area TEO) that seems to be responsible for linking the object reference frame (in TE) to the retinal reference frame (in striate and extrastriate areas; Iwai & Mishkin, 1969; Weiskrantz, 1990). Lesions in the lower part of IT (TE) lead to lasting deficits in the learning of new visual discriminations and the recollection of previously learned memories (Chow, 1954; Mishkin & Pribam, 1954; Mishkin, 1954, 1966, 1972, 1982; Iwai & Mishkin, 1969; Gross, 1973a,b; Review by Dean, 1976) while lesions of intermediate temporal cortex (TEO) impair the visual discrimination of objects per se whether the discriminanda differ in color, orientation, brightness, pattern, or shape (Gross, 1973a,b; Dean, 1976, 1982; Wilson, 1978; Gross, Bruce, Desimone, Fleming, & Gatass, 1981; Ungerleider, & Mishkin, 1982; Mishkin, Ungerleider, & Macko, 1983). As the anterior portion can be also separated in an anterior half (area TEa) more strongly linked with the limbic structures than is the posterior half (area TEp), it is thus possible to consider the ventral cortical pathway, which processes visual discrimination in primates, to include at least the following steps: V1 → V2 → V4 → TEO → TEp → TEa (Fig. 1A) (Mishkin, Ungerleider, & Macko, 1983; Yaginuma, 1990).

The dorsal pathway itself can be subdivided into two streams: (1) a dorsomedial stream involving the parieto-occipital visual area (PO) and area PG (or 7a) in the inferior parietal lobule (IPL) and (2) a dorsolateral stream involving areas MT and MST as well as the lateral intraparietal areas (Gatass, Rosa, Sousa, Piñon, Fiorani, Neuenschwander, Moura, Abrahão, & Saraiva, 1990). Neurophysiological explorations of the different parietal areas (Mountcastle, Lynch, Georgopoulos, Sakata & Acuna, 1975; Lynch, Mountcastle, Talbot, & Yin, 1977; Hyvarinen & Poranen, 1974; Sakata, Shibutani, & Kawano, 1983; Sakata, Shibutani, Kawano, & Harrington, 1985; Sakata, Shibutani, Ito, & Tsurugai, 1986) have shown that they process complementary aspects of space. The dorsolateral pathway processes visual motion (MT), with more and more integrated aspects (MST). The dorsomedial pathway seems to be involved in coding the location of visual stimuli for spatial orientation, spatial perception (Ratcliff & Davies-Jones, 1972; Lamotte & Acuna, 1978), and visual guidance of hand movements (Pohl, 1973; Buchbinder, Dixon, Hwang, May, & Glickstein, 1980; Ungerleider & Mishkin, 1982; Weiskrantz & Saunders, 1984). For example, in area 7a, neurons can combine information on the retinal location of visual stimuli and on the position of the eyes in the orbits possibly to locate visual targets in a head-centered space (Andersen & Mountcastle, 1983; Andersen, Essick, & Siegel, 1985). The two subsystems can contribute to the control of eye movements: MST neurons for smooth pursuit, 7a neurons for fixation, and area LIP for programming saccadic eye movements (neurons can hold in short-term memory the metrics of planned eye movements in motor coordinates) (Hyvarinen, 1981; Andersen, Essick, & Siegel, 1985; Andersen, Bracewell, Barash, Gnadt, & Fogassi, 1990; Gnadt & Andersen, 1986, 1988).

The dorsolateral pathway for motion perception, after V1 and V2 areas, passes via intermediate areas such as MT and MST, to the intraparietal regions while the dorsomedial pathway for spatial perception shares the two first maps V1 and V2 and then passes, through PO, POa / 7a (or PG), to the posterior parietal cortex.

From Retinal to Object-Centered Reference Frames

The global architecture of the Y-like multilayered network is organized in six maps along four sequential processing stages that model the main functional steps of the actual cortical areas (Fig. 1B): two low-level stages learning-independent and two high-level learning dependent: (1) primary areas perform low-level processing, such as elementary feature extraction, on the retinal information; (2) secondary areas perform more elaborated processing using larger RF but still with a retinal reference frame; (3) associative intermediate areas relate the resulting pattern of activity with the (4) highest level steps, which store the learned prototypes.

It is precisely *those four functional stages that* we have implemented in four processing maps. Their correspondence with actual cortical areas will be discussed further (discussion section). The two first "low-level" maps, common to the two streams, are called the primary sensory area (PSA) and the secondary sensory area (SSA). The two "high-level" maps are called the temporal associative area (tAA) and the temporal semantic area (tSA) in the "What" branch and the parietal associative area (pAA) and the parietal semantic area (pSA) in the "Where" branch.

The processing pathway leading to the parietal lobe can process both moving and static patterns, respectively, in the dorsolateral and dorsoventral streams. Given that we deal only with static pattern recognition, the "Where" branch only models the dorsomedial parietal stream.

Central-Periphery/Receptive Field Size

Visual Field Representation and Receptive Fields Size

The segregation of the visual system in two separate processing pathways matches well with the old idea of focal vision vs. ambient vision, exposed by Trevarthen (1968). The focal vision system is involved in object manipulation tasks and form recognition. It has a very high degree of acuity and concerns mainly the central visual field (foveal zone). The ambient vision system concerns the entire visual field, has a very poor degree of spatial resolution, but is particularly selective for movement and position in space of global objects. The

differences in the emphasis on the representation of central (or focal) vs. peripheral (or ambient) vision in the distinct visual areas are likely to be related to the functional division in ventral and dorsal pathways. Indeed, in the areas of the ventral stream (V4, TEO) there is emphasis on the representation of central vision while the dorsal stream has a representation either of the whole visual field or with emphasis on the periphery (in PO virtually no receptive field centers were found below 20° of eccentricity).

The second important parameter is the size of the receptive fields (RF) in the different areas of the two streams. As already outlined by Maunsell and Newsome (1987), RF size provides a good index of how the sensory information is integrated through the successive processing stages. There exists a relation between neuronal receptive field size and its eccentricity (here measured by the center of the receptive field and not to the maximal activity in the field) and we have plotted this relation (RF diameter /ecc. function) for the known cortical areas belonging to ventral and dorsal streams in Figure 2.

Three points can be made about the organization of the RF in the different maps:

1. Within the same area, the receptive field size is not constant but increases with eccentricity (in a linear fashion, at least in the lower level areas) (Gatass & Gross, 1981; Albright & Desimone, 1987; Gatass, Sousa & Rosa, 1987; Rosa, Sousa, & Gatass, 1988; Gatass, Sousa, & Gross, 1988).

2. When comparing different areas of each functional stream at a similar eccentricity value, the mean receptive field size increases significantly from one step to the next one, that is, for example from V1 to V2, from V2 to V4, from V2 to PO, resulting in a progressive integration of more and more sensory information (Hubel & Wiesel, 1962, 1965, 1968; Van Essen & Maunsell, 1983; Maunsell & Newsome, 1987; Zeki & Shipp, 1988).

3. Along the temporal pathway, the receptive field size increases gradually through the successive cortical areas but does not exceed 20° × 20° at the highest level (IT). In contrast, the receptive field size increases more rapidly along the parietal pathway. This effect is enhanced by the emphasis on the peripheral representation in this stream [cf.(1)]. At the most integrated level (PP), it can reach half or even the entire contralateral visual field, often including part of the ipsilateral one (Motter & Mountcastle, 1981; Sakata, Shibutani, & Kawano, 1983, Sakata, Shibutani, Kawano, & Harrington, 1985; Steinmetz, Motter, Dufy, & Mountcastle, 1987; Motter, Steinmetz, Dufy, & Mountcastle, 1987).

Functional Differentiation of the Two Branches of the Network

In the model, *taking into account that there is not an absolute correspondence between the processing steps and the actual cortical maps,* we have simplified the RF size of each processing step considering it as constant, and we have focused on two principles of the biological architecture: (1) the RF size increases from one step to the next one in each stream, and (2) the mean RF size increases more rapidly in the "Where" branch than in the "What" branch (in relation with the emphasis on the peripheral zone).

The distribution of RF sizes in the model reflects an important functional consequence of the RF size organization in the visual system. To learn a visual pattern, the network has to associate a configuration of local features with a particular prototype. The RF sizes affect the precision of the configuration that can be learned. In the temporal pathway, the slow increase of the RF size should maintain a precise configuration until the posterior IT (where RF size are smaller than 7°) while in anterior IT prototypes seem to be represented with receptive fields larger than 20°. We can thus hypothesize that a fundamental step in learning a configuration of local features should then occur between posterior IT and anterior IT. We have modeled this step in the relation between the associative map and the semantic map.

The transition between the successive areas is different in the two functional streams: a moderate convergence, as in the temporal pathway, allows the relative spatial

Figure 2. *Receptive fields size as a function of eccentricity in visual areas.* Neuronal receptive field sizes as a function of eccentricity (RF/ ecc function) of cortical areas in ventral (*V4, IT*), dorsomedial (PO, PP) and dorsolateral (MT, MST) streams as compared to primary and secondary visual areas (**V1, V2**), from Gatass et al. (1987) for V1, Rosa et al. (1988) for V2, Gatass et al. (1988) for V4, Desimone and Ungerleider (1986) for MT/MTp and MST, and Neuenschwander et al. (1990) for PO. For IT (dashed line), the functions are approximated from Desimone and Gross (1979) and Desimone and Ungerleider (1986), with properties similar to those of area FST, whose receptive fields are large enough to encompass the center of gaze. For PP (dashed line), approximation from Steinmetz et al. (1987), Motter et al. (1987), and Motter and Mountcastle (1981), with very large RF, up to 100° in diameter. Note that slope increases in the successive processing steps in both ventral and dorsal streams and the representation of central vision is emphasized in the ventral stream while periphery is more represented in the dorsal stream (in PO virtually no receptive field centers were found below 20° of eccentricity). This property, associated with the increase of RF size with eccentricity, results in larger RFs in the dorsal stream.

arrangements of local features to be discriminated whereas a steeper one, as in the parietal pathway, can allow the detection of global coarse patterns all over the visual field. We consider in the model this differential increase of RF size as a preset anatomical parameter and we have fixed the ratio "Where" RF/"What" RF at 9/4.

Information Coding in the Two Pathways

Neuronal Response Properties in the Two Visual Pathways

As for their receptive field sizes, the stimulus selectivity of neurons differs in the two streams and in the successive steps within each stream.

At the primary levels of processing, the ventral pathway contains a higher proportion of neurons that are mainly selective to orientation and thus more related to physical properties of patterns while neurons in the dorsomedial pathway are more selective to the direction of movement of the pattern in the visual field (Maunsell & Newsome, 1987; De Yoe & Van Essen, 1988). Orientation sensitivity in the dorsolateral stream may be responsible for processing location of stationary stimuli in space.

In the posterior parietal cortex, neurons are typically not sensory or motor in a strict sense but there appears to exist a spectrum of cells ranging from those that are activated by light stimuli and unaffected by eye movements through those with combined properties ("light-sensitive" neurons influenced by saccadic behavior) to others active during saccades but insensitive to light stimuli (oculomotor neurons or "saccade neurons"). This gradient change in functional properties suggests the possibility that such cells may be arranged in a sequential chain from visual neurons to motor ones (Motter & Mountcastle, 1981). The lateral intraparietal area neurons respond to light stimuli and have also saccade-related responses, which are mainly *pre*saccadic, while the area LIP appears to play a role in the processing of saccadic eye movements rather in motor coordinates (Barash, Andersen, Bracewell, Gnadt, & Fogassi, 1988; Barash, Bracewell, Fogassi, & Andersen, 1989; Andersen, Essick, & Siegel, 1985; Andersen, 1987, 1989).

In the temporal lobe, neurons can respond to simple stimuli, such as white and colored bars, but can also respond to much more complex stimuli including real objects such as hands and faces (Gross, Bender, & Rocha-Miranda, 1969; Gross, Rocha-Miranda, & Bender, 1972; Gross, 1973a,b; Rocha-Miranda, Bender, Gross, & Mishkin, 1975; Gross, Bender, & Mishkin, 1977; Rolls, Judge, & Sanghera, 1977; Desimone & Gross, 1979; Bruce, Desimone & Gross, 1981; Perret, Rolls, & Caan, 1982; Desimone, Albright, Gross, & Bruce, 1984). Responses to more complex stimuli increase with a posterior to anterior gradient in the ventral stream. In the posterior IT (TEO and TEp) most cells have small receptive fields (<7°), and show selective responses for simple features

such as orientation, color, and/or size of the stimuli (as in V1, V2, and V4). In the anterior IT, the proportion of these "primary" neurons falls from 72 to 9%, while the proportion of "elaborated" cells, with larger receptive fields (>10°) and optimal responses to complex stimuli (group of patterns of particular shapes), increases from 9 to 44% (Tanaka, Saito, Fukada, & Moriya, 1990). In anterior IT, elaborate responses are invariant irrespective of stimulus size, orientation or color (Miyashita, 1990). It has been demonstrated that elaborate response of cells to complex stimuli can be acquired through learning (Miyashita, 1990). Consequently they could represent the learned "prototypes" and their invariant properties should be built in the posterior IT (TEp) (Weiskrantz, 1990). The coding of "prototypes" in the anterior IT seems to correspond neither to the idea of "local" coding (grandmother cell) nor to a complete "distributed representation," but rather to an intermediate "semidistributed coding" (Tanaka, Saito, Fukada, & Moriya, 1990). Cells with different levels of complexity are intermingled among clusters that appear to associate elaborate neurons (for example, the image of the hand) and primary neurons representing partial features (for example, spatial frequency of fingers) of the optimal patterns activating the close elaborate cells (Tanaka, Saito, Fukada, & Moriya, 1990).

Local Coding of Features in Primary Areas, Semidistributed Coding of Prototypes, and Population Coding of Locations

In the model, all the units in a given map perform the same type of local processing on the afferent information flow. They are classified according to three kinds of indice (which correspond to the dimensions of the processing maps in Fig. 1B): (1) topographic (X,Y), (2) local indice (i), and (3) global indice (p).

1. The *topographic index* defines the ordered relations between the successive maps: two neighboring units in a given map have two neighboring receptive fields in the afferent map. This topography becomes less precise along the processing hierarchy. This index corresponds to the two planar coordinates (X,Y) labeling the relative location of each unit in its respective map and consequently the location of its retinal receptive field.

In the present implementation, $X=Y=16$ in PSA, $X=Y=8$ in SSA, $X=Y=4$ in the associative areas, and $X=Y=2$ in the semantic areas.

2. The *local feedforward index* (i) labels the specific filtering process that each unit performs on the sensory information. As we were interested in *static pattern processing,* and as the *outline shape* of such patterns appears to be a particularly salient feature (Gross, Rocha-Miranda, & Bender, 1972; Rocha-Miranda, Bender, Gross, & Mishkin, 1975; Gross, Bender, & Mishkin, 1977; Sato, Kawa-

mura, & Iwai, 1980; Desimone, Albright, Gross, & Bruce, 1984), we focused on *orientation selectivity* (a characteristic common to the two pathways, even if more preponderant in the temporal one). For each topographic indice (X,Y), four units $(i=4)$ of the primary sensory area (PSA) detect four orientations of edges modeling the tuning properties of the cortical columns in striate and extrastriate cortices: horizontal (H), vertical (V), and the two diagonals (D1, D2). Units of the secondary sensory area (SSA) detect six different combinations of orientations $(i=6)$ with larger receptive fields: (HH), (VV), (D1D1), (D2D2), (HV), and (D1D2). Such extraction evokes simplified models of complex and hypercomplex cells found in the striate and extrastriate cortices. This organization with 3-indices (represented by a 3D structure in Fig. 2) is maintained in the associative areas.

3. The global feedback index (p) defines a "fourth dimension" (only in the associative maps, tAA and pAA) that characterizes units receiving feedback links from groups coding a given prototype (p) in the semantic maps. These feedback links will allow us to learn to associate a configuration of local features (i, through feedforward links) with a given prototype (p, through feedback links) within the associative map. After learning, a pattern presented on the retina will activate a configuration of local features (i) and this ascending information will be matched with the stored internal representations (see Mishkin, Ungerleider, & Macko, 1983 for a neurobiological discussion).

4. The first three dimensions (X, Y, i) are maintained in groups of neurons coding for different prototypes (p) in the semantic maps (semidistributed coding) but with still larger RF sizes (less precision on X and Y). In these maps, a pattern (for example, a particular view of a given prototype) is thus not encoded in a single unit but is encoded by a population (see Desimone et al. 1984, for biological sources), i.e., is reflected in the pattern of activity across this population of units coding for different relative positions, orientations and other characteristics of the same prototype. This coding is consistent with experimental results in the inferotemporal cortex (Desimone, Albright, Gross, & Bruce, 1984; Tanaka, Saito, Fukada & Moriya, 1990). The summed population activity will be the same for various views of a prototype, even if at the level of the single cells (or automata) the pattern of activation can be different and reflects the particular characteristics of each view. Recognition of a prototype corresponds to a strong population activity in the corresponding cell cluster, and if several groups are simultaneously active, the discrimination is made through a winner-take-all mechanism.

In the present implementation, the neural network model learns 28 different prototypes which are stored in 28 different clusters of processing units. These 28 prototypes can correspond to many different incoming patterns, which are coded by different configurations of

activities within the corresponding clusters leading, however, to the same population activity. Considering that a limited number of prototypes exist [no more than 100,000 have been described (Thorpe, 1989)], such a semidistributed coding is biologically plausible.

The response in the "Where" branch is also the result of a population code. The different units vote for a given position (upper left/upper right/lower left/lower right) with a strength proportional to their activation. Since they model population of neurons in the parietal stream that project to structures commanding eye movements, a vectorial sum of the different activities can be interpreted as a shift command to reset a pattern on the fovea.

It is important to note that even if the two semantic areas express a decision about either form or position, the branch that extracts the position includes a coarse coding of the form and reciprocally the branch that extracts the form includes a coarse coding of the position.

Processing Unit

The Columnar Organization of the Cerebral Cortex

The cortical maps are not made of a uniform lattice of neurons, but the cerebral cortex is made of local circuits of cooperative neurons (pyramidal neurons and interneurons) arranged throughout the depth of the six cortical layers. This group of highly interconnected cells shares the same set of inputs and outputs and show similar selectivity to external stimuli (Szentagothai, 1975; Evarts & Tanji, 1974; Hubel & Wiesel, 1977; Mountcastle, 1978; Jones, 1981). These stereotyped neuronal circuits (about 100 neurons) are repeated, with cytoarchitectonic variations, throughout the cortical sheet (Szentagothai, 1975; Mountcastle, 1978).

The cortical layers correspond to a differential distribution of neuronal types in local circuits as well as subsets of inputs and outputs. The intermediate layer IV (which contains granular cells) divides two subsets of pyramidal neurons, the upper division (supragranular cortical layers I, II, and III) and the lower division (infragranular cortical layers V and VI). These divisions process different types of information. The granular layer receives the main sensory inputs, either directly from the thalamus (LGN and thus from the retina) or from columns of cortical areas involved in earlier stages of sensory processing (feedforward connections). The supragranular layers are mainly specialized in corticocortical connections, toward adjacent cortical zones or toward more distant cortical areas (Szentagothai, 1975; Jones, 1981; Feldman, 1984). The infragranular layers project outside the cortex toward other neural structures such as the superior colliculus or control the ascending information flow, through feedback connections (Van Essen & Maunsell, 1983; Zeki & Shipp, 1988).

These local multineuronal circuits form "cortical columns" that have been described as functional units in

the different cortical areas (Szentagothai, 1975; Hubel & Wiesel, 1977; Mountcastle, 1978; Gilbert & Wiesel, 1981). They can be labeled by different biological markers and physiological explorations that show regular patterns in surface views of the cortex (Hubel & Wiesel, 1977). In fact, different superimposed patterns may appear and could represent the overlay of different information flows (Swindale, 1990) that are combined by the local cortical interneuronal circuits.

The Learning and Activation Rules of the Processing Units Modelling the Cortical Column

The Processing Unit. The processing unit used in the network is thus not a neuron but models in a simple way this local neuronal circuit that combines the input information arriving in the different cortical layers, with different spatial arrangements (Fig. 8, explained in the Formalism section).

This model of "cortical column" (viewed as an interneuronal circuit) is organized in three major input–output divisions (Burnod, 1988; Alexandre, Burnod, Guyot, & Haton, 1988; Alexandre, Guyot, Haton, & Burnod, 1991): the upper, the intermediate, and the lower divisions in which three kinds of inputs are dispatched and produce by specific combinations two different outputs.

The processing unit combines three types of inputs:

1. *feedforward* (or bottom-up) inputs, in the intermediate division (modeling cortical granular layer IV), from previous stage units, providing sensory information from the "outside world" that has already been subject to selective filtering through previous processing steps;
2. *feedback* inputs from higher stage units (modeling top-down effects), in the upper division, which control the ascending flow of information by a gating or an inhibitory effect which is determined by both previous learning and the internal state of expectation of the system;
3. *lateral* inputs provided by neighboring units, in the upper division, which can become either excitatory or inhibitory, depending on the differential activities of the units, modeling the intracortical balance between columns.

The processing unit produces two kinds of outputs:

1. The upper division (modeling the supragranular cortical layers) makes local connections with surrounding units of the same map and long-range connections with other maps.
2. The lower division (infragranular cortical layers) projects outside the network to other structures involved in oculomotor commands or generates the feedback connections for selective control of the incoming information provided by lower maps.

We consider that interactions between feedforward and feedback connections are important for learning. We have thus implemented reciprocal connections between the associative map and the semantic map in both "What" and "Where" pathways (i.e., between tAA and tSA and between pAA and pSA). Connections between RETINA and PSA, between PSA and SSA, and between SSA and the two associative areas (tAA and pAA) involve only feedforward connections and are thus not influenced by learning.

Activity Levels. The activity range of these local cortical circuits are simplified with three different levels of activation, which correspond to three different functional states, E0, E1, and E2:

1. The null state (E0) is the result of inhibitory processes.
2. The low state (E1) models a moderate activation of the intracortical circuit, corresponding to selective attention or anticipation of a possible goal. It can be interpreted as an "hypothesis" that can dynamically propagate through the network searching for a validation.
3. The high state (E2) models a strong activity corresponding to a certainty in the detection of a specific sensory input or to the validation of an "hypothesis." Such states are typically observed during sensorimotor interactions with the external world.

Activation Rules. Activation and learning rules are detailed in the section Formalism. They are constrained by neurobiological knowledge (bottom-up constraints) and are in correspondence with elementary human-like type processing (top-down constraints) (Alexandre, Guyot, Haton, & Burnod, 1991)

The different inputs enter the unit through separate divisions (as in cortical layers). Within each division, a convolution and a thresholding are performed on subsets of inputs (corresponding to the receptive field of this unit division). The results (or global inputs) of the different divisions are integrated to compute the global output activity of the unit, taking into account the previous state of the global unit (short-term memory) and the learning coefficients (long-term memory) specific to each division. A *Strong* input can produce an excitatory or inhibitory influence depending upon the previous state of the unit. A *Moderate* input may have different influences in function of learning: it can either trigger strong activity in the target units or gate the effect of another input (as an "AND" operation) or inhibit the target units. In the present implementation, only the feedback input is dependent upon learning.

The output computation in the "What" semantic map (tSA) corresponds to a matching process between the feedforward incoming distributed activity and the learning coefficients, for both excited and inhibited values, corresponding to the stored prototypes. This dual match

operation enables two embedded patterns (such as "P" and "R") to be differentiated.

Learning Rules. The processing unit has specific learning coefficients for each division. According to these learning coefficients, each type of input may have three different effects, which depends upon the probability of input–output coactivations: (1) triggering effect, (2) gating effect, and (3) inhibitory effect. This learning can occur in one step due to weight modifications that are based on conditional probabilities of strong input–output events (see Formalism section): even with a single trial the network produces a learned response (the conditional probabilities are equal to one or zero). This response will be further refined through the following trials. This type of learning seems more realistic than progressive adjustments of synaptic weights, which need a high number of iterating steps on a given pattern or a high number of presentations of the same learning set of patterns.

EXPECTED BEHAVIOR AND PERFORMANCE OF THE NETWORK

This biologically constrained model has been tested on a set of 28 prototypes to investigate its IVR capabilities. The results match the five expectations presented in the introduction.

Learning Session: Immediate Learning

The learning corpus is constituted of a set of 28 binary patterns digitized in a 12×12 grid (the 26 capital letters of the alphabet plus two geometric figures: triangle and diamond).

During the learning session, a given pattern is presented *once* on the *central part of the retina*. At the same time, all the units in both semantic areas are inhibited (E0) except those corresponding to the adequate prototype [uniform population activation over all the local indices (*i*) of the prototype (*p*)] and its location (uniform activation corresponding to a centered position), which are forced to the stable state E2. Information propagates in the whole network according to the connectivity and the functional rules both in the bottom-up direction (through "feedforward" connections) from the retina to the two associative areas and in the top-down direction (through "feedback" connections) from semantic areas down to associative areas. At the level of the associative areas, the two flows interact and build up by learning the excitatory or inhibitory learning coefficients, which correspond to the internal representations of the external pattern. In accordance with the differences in the RF size between the two branches of the network, *two different schemes of internal representations* (Fig. 3) are obtained at the end of the learning session.
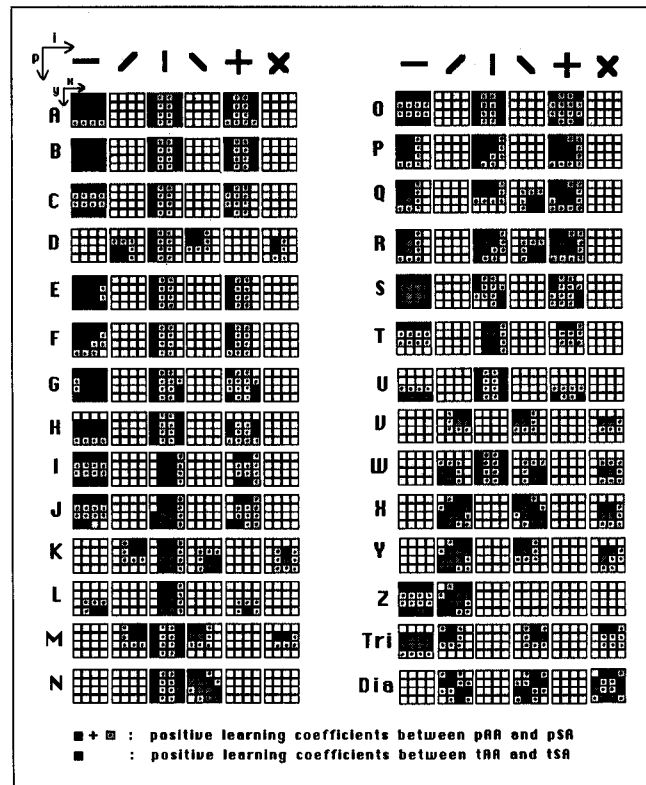
Figure 3. *Learning coefficients in temporal and parietal associative areas.* This figure compares the two sets of learning coefficients stored in the feedback connections from the semantic areas down to the corresponding associative ones: they store two "internal representations" (represented within the same map: dark squares only for the temporal coefficients and dark squares plus gray ones for the parietal ones) of the 28 learned prototypes. Each prototype (*p*) is coded in a cluster of 4(*X* × 4(*Y*) × 6(*i*) processing units (semidistributed coding, with 6 blocks of 16 units for 6 features HH, D1D1, VV, D2D2, HV, D1D2 in 16 positions). Black squares signal for both (*i*) the basic features componing each learned prototype and their topographical location (*X,Y*). Note that an adequate set of features is extracted in both the "What" and "Where" pathways, but their relative positions (spatial configuration) are maintained only in the temporal associative map (black squares).

1. In the tAA, the model has associated each prototype with a set of local features with their relative spatial arrangement. This maintained configuration of local features will allow patterns to be differentiated during the recognition session.

2. In the pAA, although the local features of each pattern have been detected as well, their relative positions have been mixed by wider RF: the local features have been spread and thus learned everywhere in the associative map. Ambiguous patterns built up with the same local features but in a different spatial configuration (like a square and the letter H) will thus be confused. However, they will always be localized, whatever their position on the retina.

To investigate the capabilities the model had acquired, after this single learning session, in pattern recognition,

all activity-dependent modifications are stopped and the different learned patterns are presented to the network with varying sizes (just one has been learned) and varying locations within the visual field (only the centered position has been learned).

Immediate Recognition of Centered Patterns

During the recognition session, the network reaches its stable state in one time step [what we consider as one time step is the time needed by the activity to propagate in a feedforward mode from the input layer (the retina) to the output ones (the two semantic areas) as defined in backpropagation algorithms] and gives a response in the two semantic areas.

Figure 4 shows that the model has a recognition rate of 100% when the patterns are presented in the same conditions as in the learning set (size 12, at the center).

We have thus performed three different sessions of tests:

1. The pattern remains at the *center* of RETINA (which is the learned position) but *varies in size.*
2. The pattern remains at the size that has been learned but *varies its location* on the retina.
3. *Both size and location change.*

Moderate Size Invariant Properties in the Temporal Branch

The pattern, *centered* on the RETINA (which is the learned position), *varies in size,* from a maximal size occupying the whole associative areas (size 13=26×26 pixels) to a minimal one which corresponds to one RF size in the primary sensory area [only seen by 9 processing units (size 3=6×6 pixels)].

Figure 4 shows that, when the size is decreased by 25% (from size 13 down to size 9), the network also

responds with a very satisfactory rate (superior to 70%) for varying sizes. Other investigations using smaller sizes for the learning set show that the recognition rates for patterns with increasing or decreasing sizes are roughly symmetrical with respect to the learned one.

What it is important to stress is the fact that the patterns within this range of sizes (i.e., from size 13 to size 9) produce patterns of activity in the tAA that perfectly match the stored internal representations (represented in Fig. 3A). This is mostly due to the RF width. This is not the case for smaller patterns whose feedforward pattern of activation concern more central units in such a way that a mismatch does occur: in fact, there is a match with a prototype that is not the good one (for example, "C" when too small is interpreted as "E").

Moderate Shift Invariant Properties without Ocular Movement

In this case, the pattern, with the learned size (size 12=24×24 pixels) *varies in location,* from one side to the other of the retina, on both the vertical and horizontal axes (corresponding to a shift up to 6 "pixels" on each axis and in the two directions).

In the case of varying location, the model shows (Fig. 5) a very interesting invariant recognition rate: it remains superior to 70% for a translation on RETINA corresponding to the size of the receptive fields of the temporal associative area ($|\delta x|$=4 and $|\delta y|$=4 "pixels" on RETINA). This is also due to the convergent connectivity.
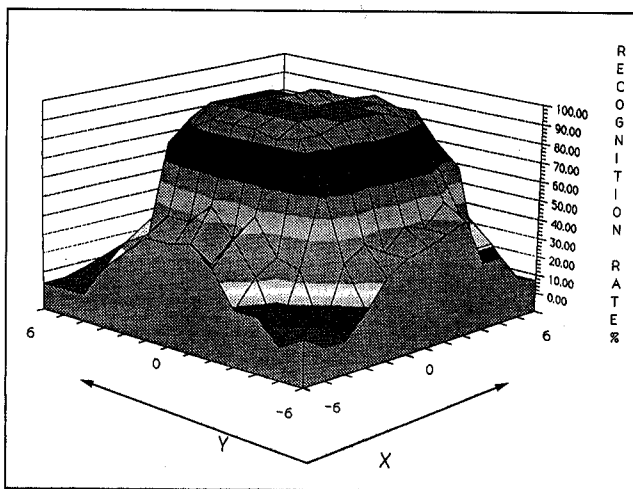

Figure 5. *Performances of the temporal pathway: effect of changing position only.* With the same learning conditions as in Figure 4 [size 12 and central position, position (0,0) in this figure], the 28 patterns are presented, during the recognition session, in the same size but with varying positions, from one side to the other of RETINA [given the size of the retina (32 × 32) and the size of the pattern (24 × 24), its position is changed up to 6 pixels in both the X and Y axes). Note that the recognition rate remains superior to 70% for a shift up to 4 pixels in the two directions and along the X and Y axes. This shift globally corresponds to the surface of the retina (RF size) "seen" by the processing units of the temporal associative area.

Figure 4. *Performances of the temporal pathway: effect of changing size only.* After the 28 patterns have been learned at one size (size 12, 24 × 24 pixels) and in the central position, they are presented, during the recognition session, in the central position but with varying sizes: from a size 13 (26 × 26 pixels) down to a size 3 (9 × 9 pixels). Note that the recognition rate remains above 70% despite decreases from the learned size of up to 33%.

Parietotemporal Cooperation for Invariant Recognition

We focus on the conditions where the "What" and the "Where" branches tightly cooperate to perform successful pattern recognition.

Figure 6 shows the behavior of the Y-like network *when both size and location* of the presented pattern *are changed* [for each different size (13 to 3), the pattern is presented in *all the locations* over the retina].

When the pattern is not too small, the "What" branch can perform direct recognition at different locations, mainly around the center of RETINA: the positions of the patterns where this direct recognition is satisfactory are represented in Figure 6 by the darkened squares. In this case ("temporal satisfaction"), the "where" computation

Recognition

Ocular command
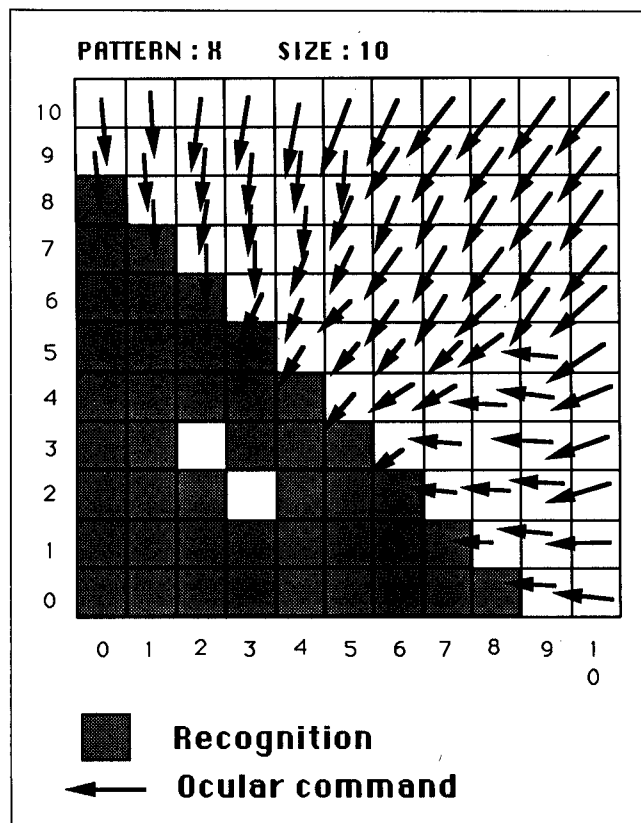
Figure 6. *Cooperation between temporal and parietal pathways.* With the same learning paradigm as Figures 4 and 5, a pattern is presented during the recognition session with varying sizes and position shifts on the retina of up to −10/+10 "pixels." This figure shows the behavior of the double-branched network for a prototype (here the capital letter "X") presented with one size (size 10, 20 × 20 pixels) in all the positions (in the upper quarter of the retina, the other quarters can be deduced by symmetry). Dark squares represent the positions where the pattern is directly recognized by the "What" branch. Arrows represent the vectorial population coding of an ocular movement performed in the "Where" branch when the "What" branch fails to recognize the pattern. Notice that these vectors contain the directional information needed to reset the pattern closer to the positions where it is directly recognized (dark squares). Furthermore, the amplitude of the vector increases with the distance of the pattern from the center.

is performed in parallel in the parietal branch but is not used for recognition. A recognized pattern finally results in a well-contrasted activity (E0, E2) in the tSA. We consider that the overall effect of this "temporal" activity is inhibitory on the output from the "Where" branch (driving the ocular command).

By contrast, failure of recognition corresponds to spread of activities (E0, E1 and E2) within the tSA, without any emergent well-contrasted pattern. The overall effect on the "Where" branch is considered as an excitatory gating effect (a "call for help") on its oculomotor output. The "Where" branch has already extracted a global location of the pattern displayed on the retina (population coding). The output effect of the parietal areas toward the structures that command the extraocular muscles (colliculus, FEF) is modeled by a vector: the population code is transformed into a motor command (vectors shown in Fig. 6), which drives the movement to be performed to reset the pattern in the foveal region.

Figure 6 shows that the "Where" branch not only computes the *good direction,* from the target location toward the center of the visual field, but also calculates an *amplitude,* which increases, in a step-like fashion, with the eccentricity of the target: the greater the eccentricity of the target, the larger the amplitude of the resetting vector.

The motor command positions the pattern closer to the center of the fovea and a few time steps are sufficient to reset the pattern in the region where the "What" branch recognizes it. This resetting operation can fail for smaller patterns: in this case, it is necessary either to change the internal representation ("internal zoom," exposed in the Discussion section) or to move toward the object ("external zoom"): these two adaptive processes can be triggered by the persisting spread activity ("lack of satisfaction") in the "What" branch.

Furthermore, it is important to note that the response given by the "Where" branch is linked to the "parietal prerecognition" of a possible pattern: the "Where" branch indeed will not localize a random signal that does not have a similar association of local features to the previously learned patterns (as a "list" and not as a "configuration").

DISCUSSION

In this article we have outlined a computational solution to the invariant recognition problem—invariance in size, invariance in position—which is based on neurobiological knowledge. The invariance property in the whole visual field mainly comes from the cooperation between two regions that are a priori known to extract two different types of information, one that has limited invariant capacities for object recognition in the center of the visual field and the other that can extract object locations in the periphery and drive eye movements to reset the pattern in the central region. A Y-like double branched network (constraint 1) allows us to extract these two

different types of information, form and location, with an optimal tradeoff between precision of recognition and invariant capabilities in each branch, which is due to a differential distribution of RF sizes and visual field representation in the two streams (constraint 3). Invariant capabilities in the central part of the visual field are due to the increase of RF size along the successive steps (constraint 2) of the "What" pathway. These steps have clearly different functional roles since only the "associative" area has the optimal tradeoff between the size of the RFs and the maintenance of spatial configuration of features to learn patterns efficiently. A semidistributed coding of information (constraint 4) allows us to represent in the network a limited number of "prototypes" by independent neuronal groups: each neuron can optimally learn a specific feature of a given prototype and different views of the same prototype are represented by different patterns of activation within the same neuronal group. The lamination of the local cortical circuits in different input–output divisions (constraint 5) allow us to integrate information along each stream (through feedforward processing), to associate local features and prototypes in the associative areas (through feedforward and feedback interactions), and to achieve the functional cooperation between the two branches (through "lateral" interactions). The learning rules of these units are based on conditional probabilities (consequence of a multicellular circuit) that allow immediate learning (constraint 6). The four last constraints represent the expected performances of the network and this paper shows how they can be obtained when the architecture of the network and the properties of the processing units are constrained by the six first constraints.

Our proposition to explain the phenomenon of "stimulus equivalence" (Rocha-Miranda, Bender, Gross, & Mishkin, 1975; Mishkin, Ungerleider, & Macko, 1983; Gross & Mishkin, 1977; Seacord, Gross, & Mishkin 1979) for changes in location and size over the central 20° or so of the retina lies thus in anatomical constraints due to the size of the receptive fields in the associative areas. Recovering global information about "what is it everywhere it could be" is obtained due to the parallel extraction of these two types of information: when the temporal branch cannot directly discriminate an object, a shift of the pattern close to the central zone can be performed, as the result of a resetting vector computed by the parietal branch, until the temporal branch can perform a successful recognition.

This model is based on 10 constraints, five linked with the functional architecture of the cortex (bottom-up constraints) and five linked with the expected behavior of the system (top-down constraints). Our aim was not to investigate the effect of these 10 constraints considered separately since they coexist in the visual system of primates and work cooperatively at different levels. The goal of this paper is rather to show how these constraints can

be simply associated in a neural net to achieve invariant visual recognition.

Comparisons with Other Models

This model shares some particular properties with other neural networks used to perform visual pattern recognition (Rueckl, Cave & Kosslyn, 1989; Bienenstock & Von der Malsburg, 1987; Finkel & Edelman, 1990; Fukushima, 1988) but its originality lies in its commitment to support neurobiological and psychological constraints at different levels of organization.

As in other models (Fukushima, 1988; Glünder, 1990), the invariance property is first obtained by a combination of feature extractions (AND) and converging receptive fields (OR). However, in the present model, these two parameters are more directly linked to architectural properties of the visual system and are not the unique source of invariant capacities due to two complementary processes: (1) receptive fields size and (2) sensorimotor cooperation in a similar way and with similar limits as human visual recognition (for patterns learned at only one location).

As in other models (Rumelhart, Hinton, & Williams, 1986), the learning signal is provided by a feedback flow of information. But in our model, the feedback connections do not send an "abstract error signal" but rather neuronal activities representing the top-down expectations of the system. Furthermore, these feedback projections are modifiable by learning, providing a semidistributed storage of the learned prototypes that will be matched during the recognition session with the incoming patterns.

As in a recent model (Rueckl, Cave, & Kosslyn, 1989), we stress the computational advantages of the physical segregation of the "What" and the "Where" functions in two separate pathways. In their model, based on the backpropagation algorithm, they have compared two networks that both have to recognize and locate patterns in a retina, but one is fully interconnected and the other one is built with two independent hidden layers and output layers. The second architecture appears to be more efficient in terms of recognition rate and speed of learning *in each independent branch*. In our model, the two pathways are *a priori distinct* with *specific features* (RF sizes and indices) and *cooperate* for a common function, namely the invariant visual recognition.

An important characteristic of this model is the nature of the basic processing unit, which is very different from those typically used in neural networks. This unit does not model a unique neuron but rather a cortical neural circuit that can deal with various types of architectures and has specific learning and activation rules that allow a fast learning speed. To call it a "column" refers more to the vertical arrangement of these neurons in the depth

of the cortex (as stressed in the original physiological paper of Mountcastle in accordance with the classical images of Ramon y Cajal) than to the different "surface views," which have been obtained in the two last decades with different types of metabolic markers and anatomical tracers and which do not give a uniform view of the cortex as stressed by Swindale (1990) since they are linked most of the time to the specific marked or activated inputs that may have different "lateral" extensions, segregated or overlapped. The processing unit in our model is defined as a combination of several indices and would thus model a vertical circuit in the cortex intersecting several subsets of inputs and outputs that could have different spatial organizations and extensions. Such processing units are thus consistent with a wide variety of views concerning the nature of cortical columns (Swindale, 1990). Another original feature of this model is to propose a semidistributed coding of the stored prototypes which is in accordance with experimental data about the temporal lobe (Tanaka, Saito, Fukada, & Moriya, 1990).

Limitations and Potentialities of the Model

Even with these 10 constraints, we cannot ignore of course that the situation is much more complex in primate visual systems.

Almost 20 different areas have been identified on the basis of various criteria and each of them is characterized by different and complementary anatomical and functional properties. But we have made the voluntary—and reasonable—choice to focus on a minimal number of processing steps that we consider as necessary to perform invariant visual recognition. Similarly we have focused on RF size distribution within the visual pathway but it is important to point out that all the characteristics of the receptive fields are not fully known: (1) there is a scatter in the size of the RF for each eccentricity and this scatter may be of physiological importance; and (2) the characteristics of the RF around the center of the visual field are less known for higher levels areas. The size of the RF varies with the attentional state of the animal (Moran & Desimone, 1985): when the attention of an animal is focused on one of two simultaneously presented stimuli, the effect of the other one is decreased. In the model, the RF structure of the PSA units is preset to yield simple cell-like properties described in the striate cortex of the monkey. Since this organization is stabilized during a critical period that we consider finished when learning prototypes occurs, we did not make these features extractor adaptive. *The only adaptive connections in our model are feedback connections that are modifiable in either excitatory or inhibitory direction.* Lateral and long-range connections could also be adapt-

able with similar learning rules. They could introduce supplementary controls on the recognition process.

The present model has focused on temporoparietal interactions for invariant visual recognition. With the same types of units, it could include in the future other known properties of the visual system in order to deal with more natural environments:

1. Our model is restricted to 2D patterns identification. To recognize 3D patterns, it should include learning of different prototypic views.

2. The parietal coding of saccade vectors is computed in retinal coordinates and could be extended toward other coordinate systems like the head-centered reference frame (see, for example, Andersen, Essick, & Siegel, 1985; Burnod, Caminiti, Johnson, Grandguillaume, & Otto, 1990).

3. It will also be important to take into account the parallel cooperation of the colliculus for the extraction of location in space (for discussion, see Mishkin, Ungerleider, & Macko, 1983).

4. The model shows the efficiency of temporoparietal cooperation in recognizing a single pattern whatever its position in the input retina. To improve its performance with more natural visual scenes including several objects, the associative and semantic areas should contain a finer partitioning of the retinal information; to separate patterns and attentional processes would be necessary to enhance one particular pattern among the others to foveate toward it.

5. Another simple extension of the model involves generalizing the role of the parietal lobe from the location of different global targets in the peripheral visual field to the location of several fixation points *within* the same object in the pericentral zone. In both cases the parietal lobe will allow us to build, through the successive foveations, an internal representation of spatial relations either between different objects of the visual scene or between the different local features constituting a given object.

To explore a visual scene with more than one object, or to explore different parts of the same object, it should be necessary to take into account two complementary processes:

1. Sequential local enhancements of activity in the parietal region (corresponding to the points that will be successively foveated) as "attentional spots." These successive enhancements may have different origins such as local prerecognition of groupings of specific features in the parietal areas, temporal prerecognition with a coarse coding of the direction of its locus (as already coded in the model), frontal drive, given, for example, by internal expectations or an instruction about the attended location, or an internal gradient, like a Gaussian filter around

the present point of fixation, enhancing the nearest pattern with respect to the current fixation point.

2. Control by the temporal areas on eye movements until recognition is achieved.

For more complex patterns, we could hypothesize a process at the level of the parietal areas in which the two kinds of information are combined: (1) feedforward information creates a landscape of activation widely spread with several mountains and valleys and (2) lateral input coming from the temporal lobe enhances one peak (only one saccade toward a single object is performed at the same time). Step by step, the different points of the visual scene can thus be successively explored until recognition is achieved.

We hypothesize that the same kind of processing may be involved for the recognition of large-scale patterns by successive foveations: as different objects within a visual scene are successively foveated, within a given object, the different points of interest may be successively focused according both to peripheral local features and global object temporal prerecognitions.

Segregation of the Two Pathways

From a biological point of view, the segregation between the two functional pathways is still an object of discussion. This model does not exclude the existence of either a certain degree of convergent interaction between different features (for example, among Y, X, and W channels) or a certain degree of hierarchical organization between processing modules, but it does emphasize the importance of the separate projections and parallel processing. In the present model, we have considered only the direct connections within each direct pathway. The existence of additional interconnections between the two pathways and common connections from the STS region (Morel & Bullier, 1990) would correspond to functionally different inputs involved in linking and controlling the two separate direct processing pathways. Such additional inputs would not change the properties of the model since they will enter the processing units through different input layers having specific activation and learning rules, for example, for control and timing correlation.

Inputs from central vision are more important for the object recognition function of IT cortex (Mishkin & Ungerleider, 1983). The inferotemporal cortex receives a heavy projection from the "foveal prestriate cortex" (Cragg & Ainsworth, 1969; Zeki, 1969), whereas inputs from central and peripheral visual field (VF) are equally important for the visuospatial functions of the posterior parietal cortex (Mishkin, Ungerleider, & Macko, 1983). In our model, the segregation Central VF-Temporal cortex vs. Peripheral VF-Parietal cortex is functionally implemented: the temporal recognition works only in the pericentral region of the VF while the parietal action is

triggered only in the case of eccentric patterns (gating signal from the temporal branch).

"Where" Branch and Properties of the Parietal Neurons

In the posterior parietal cortex, neurons are not sensory or motor only but there is a gradient from "light-sensitive" neurons to oculomotor neurons. In the present model, we have considered the parietal semantic map units to be light sensitive and to contain, through a population code, the information necessary to drive a motor saccade toward the light target. What we suggest is that associative areas of the dorsomedial stream are capable of some pattern recognition in terms of a list of local features (no topographical organization can be learned because of the wide receptive fields). Thus it will not localize random patterns of activation but rather blurred associations of local features that could correspond to "a possible interesting pattern."

In the present implementation, we have just considered the dorsomedial pathway of the dorsal stream. We could easily generalize the properties of the dorsomedial stream to the dorsolateral one, replacing, for example, the orientations extractors by direction and speed of motion detectors. The addition of a cooperative motion pathway would be very interesting in that motion processing can help to segregate figure from ground.

"What" Branch and Temporal Neurons

Sato, Kawamura, and Iwai (1980) have described two types of IT neurons: a given neuron responsive to a global pattern may also be triggered by its components (type 1) or may not (type 2). We hypothesized that these two classes correspond to two successive processing steps. In our model, in the associative areas in which patterns are coded in a semidistributed fashion, clusters of processing units respond both to global patterns and to their local components. However, at the highest level (modeling the TEa area) a global population activity value is computed for each prototype cluster. Units respond not to the presence (or the absence) of the basic components separately (as an OR gate that operates at lower processing steps) but only to their *combined presence* (or absence, as with an AND gate). This property allows us to enhance the differentiation between patterns.

In the present model, learning of the different prototypes takes place somewhere in between an associative area whose moderate RF size allows us to keep the spatial configuration of local features (as TEp?) and a "semantic" area storing the global object in an object-centered internal representation (as TEa?). This hypothesis is consistent with data on the temporal lobe (Mishkin, 1990; Weiskrantz, 1990; Miyashita, 1990).

In TE area, there is no sign of clear visuotopic orga-

nization (Desimone & Gross, 1979). Even if we have described topographic indices X and Y in the two semantic areas, they do not organize the map at this level and appear only as "secondary" indices, repeated in different subpopulations and thus more related to pattern coding.

A Parallel Processing Focused on Different Sizes

For the different areas, RF sizes increase with eccentricity. However, a significant scattering in RF size seems to exist at different eccentricities and may have a functional meaning. It is possible to hypothesize the presence in a given area of multiple levels of representation of the retinal input. Small patterns would be processed with a high resolution and this resolution would decrease while the size of the pattern increases. Consequently, the number of processing units necessary to discriminate a configuration would thus remain constant independently of the size of the pattern. We can consider that the visual information may be processed by parallel channels with different degrees of resolution in relation to the global size of the pattern (such as an internal "zoom" effect) and that the more efficient one should perform successful recognition. These parallel channels would work within a range of minimal to maximal possible sizes: patterns that are too small would fall under the resolution level and large ones could extend beyond the perifoveal region of recognition. But, within this range, it would be possible to consider that each parallel channel corresponds to a network such as the "What" branch. The total size invariance capabilities should cumulate the capabilities of each "What" subnetwork. Outside this range, a motor command (body movement toward the object or manipulation of the object to bring it closer to face) will be necessary to optimize recognition by resetting it back to the well-working domain. Within the range of direct recognition, it is possible to compute the optimal number of subnetworks needed to cover the range.[1] Furthermore, as the learned stored configuration is the same for all scales, we can hypothesize that the learning can occur with one presentation only, in one size: learning can be stored indeed at the level of a common associative area to which the different processing channels should access. One of those channels may be selected each time, for example, in direct dependence with the maximal eccentricity of the retinal information (or the proprioceptive signal from ocular muscles responsible for convergence on the target).

Temporoparietal Interactions

Direct anatomical relations have been described between the temporal and the parietal pathways (Morel & Bullier, 1990). It is very interesting to note that (1) from the parietal to the temporal lobe, the parietal region called

POa (in direct relation with a parietal region which command eye movements) projects to TEO (intermediate region of IT, related to identification of objects per se); and (2) in the reverse direction, the anterior region of the superior temporal sulcus (STS), more related to visual recognition, projects to POa. The function of these connections is not fully understood. Modeling allows us to explore the functional role of a *direct cooperation* between "What" and "Where" pathways. The processing units can integrate such types of interactions in two stages:

1. Long-range direct connections from temporal areas to parietal ones could support the direct temporal call toward the parietal lobe: when no contrasted activity emerge within the temporal semantic area, this area can learn to gate the parietal control of an oculomotor action in order to enhance recognition through an extracortical feedback loop. We expect a strong input signaling for the global status of recognition: this signal could be provided by STS (see Morel & Bullier, 1990). The modification of learning coefficients in these long-range connections can increase the efficiency of this process. The temporal logic of these rules corresponds to operant conditioning: if a unit activated by several inputs has a strong activity level (E2) and if another of its inputs is strongly reactivated (E2), this input will gain more influence by learning. If a strong input (E2) follows an inhibited activity of the unit (E0), the input will lose influence by learning. In this case, it is important to stress that there are two different aspects of the timing of parietal activity: (1) a direct response is processed by the parietal lobe, even if there is no eye movement, with a rather short latency and (2) this first response may be enhanced by several different controls either from the temporal lobe or from other cortical structures.

2. In a second phase, an important contribution of the parietal lobe areas to recognition could be to provide the positional information directly to the temporal areas. This information could be used to shift the ascending input pattern, when displayed in perifoveal zone, *within the temporal associative area* itself, to match it with the previously stored information (learned in the central position). This *intracortical shift* could be sufficient to avoid the need for a global reset of the retinal input performed by an ocular or a body movement (Fig. 7) and then to efficiently parallelize the recognition process (Anderson & Van Essen, 1987). An additional interesting property of such a process would be the possibility that such partial shifts could be used to correct local deformations.

The processing units in this model allow such interactions to be taken into account. In the temporal associative area, units could directly combine (1) feedforward information from the retina, (2) positional information from parietal areas, and (3) feedback information signaling for success or failure of the recognition process.

The learning logic of this model predicts that learning will mostly occur when ocular movements are made to reset the pattern in the fovea (as described in the first part of this paper; operant conditioning). The information coming from the parietal pathway that can help temporal recognition will get an increased influence: this effective resetting will produce a satisfactory recognition and then a strong feedback signal will locally reinforce the previously activated parietotemporal circuit. After learning, this reinforced interaction will be sufficient to produce an anticipatory shift in the temporal pathway, improving temporal recognition without need to make an effective ocular saccade (Fig. 7).

EXPERIMENTAL PREDICTIONS OF THE MODEL

From this model, we can make five major experimental predictions.
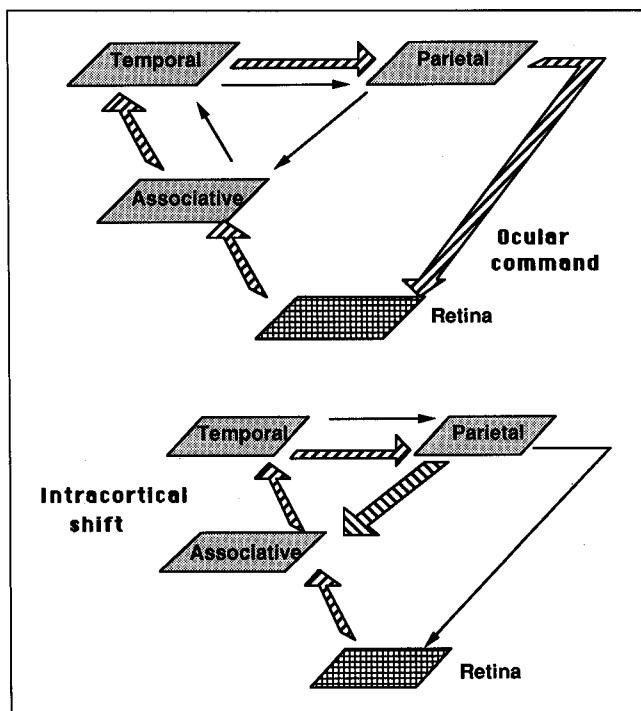
Figure 7. *Direct parietotemporal interaction to improve direct temporal recognition without ocular movements.* (Top) When it fails to recognize, the temporal areas gate the output command of parietal areas, which reset the pattern in the central region of the retina, via an *external feedback (oculomotor) loop.* We hypothesize that learning can occur in the connections from the parietal to the temporal associative area, directly associating a positional coding of the pattern in the parietal pathway before ocular movement, with a shift of the pattern representation in the temporal associative map due to the resetting ocular movement. (Bottom) After learning, a direct parietotemporal activation can shift the visual information, *directly at the level of the temporal associative area, without need of effective ocular movements.*

Psychophysical Experiments

When new patterns are learned in a given position and when ocular movements are not allowed, there are very strong limits in shift-invariance capacities. Our prediction is that these limits can be directly related to the receptive field size in an area, which is intermediate between V4 and TEa (for example TEp or TEO). We can consider that previously unknown patterns, learned in one position only, will interest in this intermediate area a small group of cells coding for the ascending information flow. There should exist a zone of shift invariance around this learned position, which should correspond to the surface covered by the receptive field of these cells. When the position crosses the border of the receptive field, in this intermediate area, the performances should drop with a curve of performances close to the one obtained by simulation and shown in Figure 4.

Asymmetrical Connectivity between the Temporal and the Parietal Lobes

Our model makes some predictions concerning temporoparietal connectivity and the corresponding interactions.

From Temporal to Parietal Lobes

The most efficient connections from the temporal lobe to the parietal lobe should be from higher temporal level (where patterns are recognized) to a region of the parietal lobe where the target position is computed and close to the regions that command ocular movements (i.e., projections either to subcortical regions as the SC or to cortical regions as LIP or the FEF). These connections should be necessary to trigger the ocular movement command required to remove ambiguities concerning the pattern to be recognized. We expect, for conveying this kind of global triggering signal, that this connection does not need to be dense or topographically organized.

From Parietal to Temporal Lobe

Conversely, the connections from the parietal to the temporal lobe need not be symmetrical to those previously described. In this case, the information about target position may be sent more efficiently to an intermediate area in the temporal stream (such as TEO or TEp) so that the resulting global information after combination could match the prototypes stored in the TEa cortex. We predict that this projection should be highly divergent so that each region of this intermediate temporal stream area can receive a positional information.

Physiological Experiments: Balance of Activity between the Temporal and Parietal Areas

As a result of the cooperative mechanism implemented in our model, there should be a "flip-flop"-type pattern of activity between temporal and parietal areas: when a pattern is recognized there is either a strong contrasting activity in a restricted part of the terminal area in the temporal pathway while the activity in the parietal terminal area remains moderate, or conversely, there is a low and widely distributed activity in the temporal area (ambiguous patterns) producing a strong enhancement (a gating) of the already activated parietal area.

Anatomy: Strong Anatomical Divergence within the Temporal Pathway

We have a priori implemented a strong divergence from SSA to tAA over the global indice (p). This divergence is essential in our model to learn each kind of prototype. We thus expect to find some intermediate temporal area (between V2 and TEa) with RF sizes allowing us to maintain the spatial configuration of the patterns and "several" repetitions of the representation of the retina (at least as much as the different prototypes).

Physiology and Attentional Process

In our model, the tAA output is gated by the feedback activity coming from the temporal semantic area corresponding to the stored learned patterns. We thus expect the propagation of activity along the temporal pathway to be strongly dependent on attentional processes. Without attention, we expect the ascending activity flow to fade away after V2 or V4 while continuing to higher levels under attentional conditions. This type of propagation could be easily tested in humans by PET studies.

FORMALISM (Fig. 8)

Notations

$E_{s/u}^v(t)$ Input from unit u to unit s in layer v

$I_s^v(t)$ global input per layer v of unit s

$A_s^v(t)$ Influence of the global input in the layer v of unit s

$A_s(t)$ Global output of the unit labeled s

Three Levels of Activity Ei ($i=1,3$)

$E_{s/u}^v(t)$, $I_s^v(t)$ $A_s(t)$ = E2   Strong activities (high firing frequency)

$E_{s/u}^v(t)$, $I_s^v(t)$, $A_s(t)$ = E1   Weak activities (low firing frequency)

$E_{s/u}^v(t)$, $I_s^v(t)$, $A_s(t)$ = E0   Null activities (inhibited)
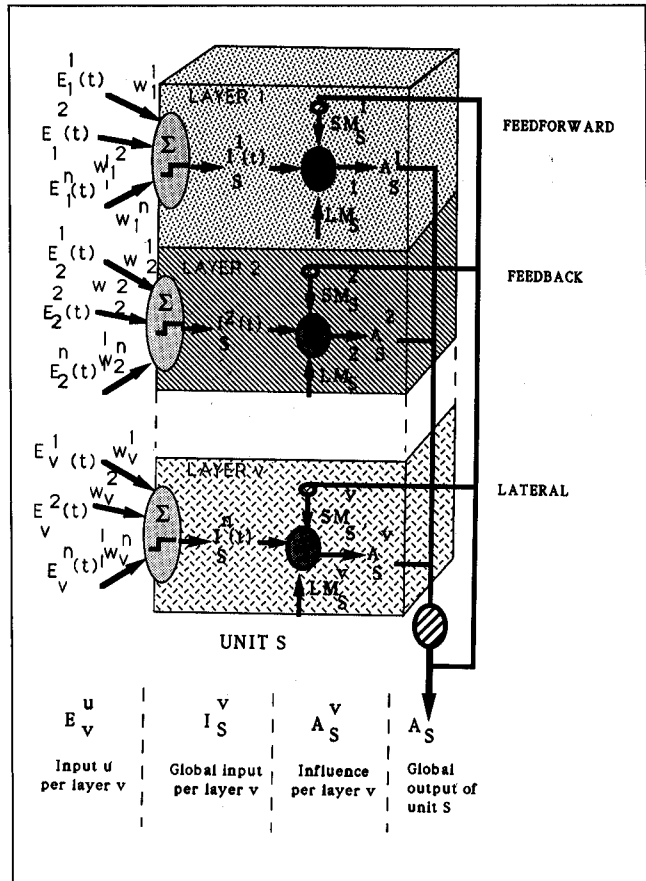

Figure 8. *Formalism.* Each unit "S" of given area receives a certain number of inputs E within each layer "v." This figure summarizes the processing steps performed on these inputs to compute the global output of the unit "S." For precisions, refer to section "FORMALISM."

Within Each Layer v

$w_{s/u}^v$ Weighting coefficients applied on the inputs u (input mask) arriving to layer v of unit s

$\mathbf{W}_s^v = \{w_{s/u}^v\}$ Vector of weighting coefficients

$\mathbf{E}_s^v(t) = \{E_{s/u}^v(t)\}$ Vector of inputs arriving in layer v

Activations Rules

Global Input per Layer v

A unit s receives, in each layer v, inputs $\mathbf{E}_s^v(t)=\{E_{s/u}^v(t)\}$ from other columns u.

Inputs $E_{s/u}^v(t)$ are convoluted with an input mask $\mathbf{W}_s^v = \{w_{s/u}^v\}$ and thresholded (function $g_{\theta,\theta'}$) to produce a "global input" per layer v: $I_s^v(t)$.

$$I_s^v(t) = g_{\theta,\theta'}\left[\sum_v w_{s/u}^v(t)\, E_{s/u}^v(t)\right] = g_{\theta,\theta'}[{}^t\mathbf{W}_s^v\, \mathbf{E}_s^v(t)]$$

with $g_{\theta,\theta'} = \delta[\theta,\theta'] + 2\delta[\theta',+\infty]$ and $\theta < \theta'$, and where θ and θ' are the two thresholds that separate the three levels of activity (given by $g_{\theta,\theta'}$).

Influence per Layer

The influence $A_s^v(t)$ on the global output (A_s) of the global input $I_s^v(t)$ depends on the global inputs $I_s^v(t)$, on a function (per layer) of the previous state of the unit ("short-term memory" factor SM_s^v) and on previous learning of layer v ("long-term memory" factor LM_s^v).

$$A_s^v(t) = SM_s^v\{^2I_s^v(t)\} + LM_s^v\{^1I_s^v(t)\}$$

where $^kI_s^v(t) = 1$ if $I_s^v(t) = Ek$ else $^kI_s^v(t) = 0$, $\quad k = 1,2$.

Global Output of the Unit

The global output is a sum of the influences per layer v:

$$A_s(t + dt) = \sum_v A_s^v(t)$$

Learning Rules

Short-Term Memory

$$SM_s^v = {}^s\alpha^v f_\theta[A_s^v(t-1)] + {}^s\beta^v f_{\theta'}[A_s^v(t-1)] + {}^s\gamma^v$$

where $\theta, \theta' \in N$, $f_\theta = \delta_{[\theta,+\infty]}$, $f_\theta = \delta_{[\theta,+\infty]}$, and $^s\alpha^v$, $^s\beta^v$, $^s\gamma^v \in N$ are parameters that are layer-specific (index v) and define short-memory effects (index s)

Long-Term Memory (indice 1)

Learning depends on conditional probability $P2_s^v$ of strong input $I_s^v(t)$ after strong outputs $A_s(t-1)$.

$$LM_s^v = {}^l\alpha^v f_\theta(P2_s^v) + {}^l\beta^v f_{\theta'}(P2_s^v) + {}^l\gamma^v$$

where $\theta, \theta' \in N$, $f_\theta = \delta_{[\theta,+\infty]}$, $f_{\theta'} = \delta_{[\theta',+\infty]}$, $\theta' > \theta$, and $^l\alpha^v$, $^l\beta^v$, $^l\gamma^v \in N$ are other parameters that are also layer-dependent (index v) and define the long-term effects (index "l"), where $P2_s^v$ represents the probability that the layer "v" has a strong global input $I_s^v(t)$ when the global output of the unit was strong at the time step before $(t - 1)$:

$$P2_s^v = \text{Prob}[A_s(t-1) = E2/I_s^v(t = E2)]$$

Value of the Different Coefficients per Layer v

Layers	$^s\alpha^v$	$^s\beta^v$	$^s\gamma^v$	$^l\alpha^v$	$^l\beta^v$	$^l\gamma^v$
Feedforward	0	0	2	0	0	1
Feedback	1	1	0	2	1	1
Local	0	0	0	0	0	0

Note

1. This computation takes into account the relative size-invariance within the prototypic subnetwork that we have already tested (Fig. 4). This relative size invariance can be represented by a factor R, which represents the ratio between the learned

size and the lower satisfactory limit. The relative pixel size between two successive subnetworks is equal to r in order to cover all the sizes between the minimal and the maximal cutoffs. The number n of subnetworks necessary to cover the whole range between the maximal scale (S_{max}) and the minimum one (S_{min}) is given by the equation $R^n = S_{max}/S_{min}$. The network can recognize all sizes if $r = R$ and then if the number of subnetworks is

$$n = (\log S_{max}/S_{min} / \log(r)$$

With this equation, it is interesting to see that the total number of subnetworks n does not increase in a linear fashion with the number of scales but with the log of this number.

REFERENCES

Albright, T. D., & Desimone, R. (1987). Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. *Experimental Brain Research, 65,* 582–592.

Alexandre, F., Burnod, Y., Guyot, F., & Haton, J. P. (1988). La colonne corticale, unité de base pour des réseaux multicouches. *Comptes Rendus de l'Académie des Sciences, Série III, 309,* 259–264.

Alexandre, F., Guyot, F., Haton, J. P., & Burnod, Y. (1991). The cortical column: A new processing unit for multilayered networks. *Neural Networks, 4,* 15–25.

Andersen, R. A. (1987). The role of the inferior parietal lobule in spatial perception and visual-motor integration. In F. Plum, V. B. Mountcastle, and S. T. Geiger (Eds.), *The handbook of physiology. Section 1: The nervous system. Vol. IV. Higher functions of the brain. Part 2* (pp. 483–518). Bethesda, MD: American Physiological Society.

Andersen, R. A. (1989). Visual and eye movement functions of the posterior parietal cortex. *Annual Review of Neuroscience, 12,* 377–403.

Andersen, R. A., Bracewell, R. M., Barash, S., Gnadt, J. W., & Fogassi, L. (1990). Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of Macaque. *Journal of Neuroscience, 10,* 1176–1196.

Andersen, R. A., Essick, G. K., & Siegel, R.M. (1985). Encoding of spatial location by posterior parietal neurons. *Science, 230,* 456–458.

Andersen, R. A., & Mountcastle, V. B. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *Journal of Neuroscience, 3,* 532–548.

Anderson, C. H., & Van Essen, D.C. (1987). Shifter circuits: A computational strategy for dynamic aspects of visual processings. *Proceedings of the National Academy of Sciences of the U.S.A., 84,* 6297–6301.

Barash, S., Andersen, R., Bracewell, M., Gnadt, J., and Fogassi, L. (1988). Saccade-related activity in area LIP. *Neuroscience Abstracts, 14,* 203.

Barash, S., Bracewell, M., Fogassi, L., & Andersen, R. (1989). Interactions of visual and motor-planning activities in the lateral intra-parietal area (LIP). *Neuroscience Abstracts, 15,* 1203.

Bienenstock, E., & Von der Malsburg, C. (1987). A neural network for invariant pattern recognition. *Europhysics Letters, 4 (1),* 121–126.

Buchbinder, S., Dixon B., Hwang, Y. W., May, J. G., & Glickstein, M. (1980). The effects of cortical lesions on visual

guidance of the hand. *Society for Neuroscience Abstracts, 6,* 675.

Burnod, Y. (1988). *An adaptive neural network: The cerebral cortex.* Paris: Masson.

Burnod, Y., Caminiti, R., Johnson, P. B., Grandguillaume, P., & Otto, I. (1990). Model of visuo-motor transformations performed by the cerebral cortex to command arm movements at visual targets in 3-D space. In Eckmiller (Ed.), *Advanced neural computers.* Amsterdam: Elsevier Science Publishers B.V.

Bruce, C. J., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the Macaque. *Journal of Neurophysiology, 46,* 369–384.

Chow, K. L. (1950). A retrograde cell degeneration study of the cortical projection field of the pulvinar in the monkey. *Journal of Comparative Neurology, 93,* 313–340.

Chow, K. L. (1954). Effects of temporal neocortical ablation on visual discrimination learning sets in monkeys. *Journal of Comparative Physiology and Psychology, 47,* 194–198.

Cowey, A., & Gross, C. G. (1970). Effects of foveal prestriate and inferotemporal lesions on visual discrimination by rhesus monkey. *Experimental. Brain Research, 11,* 128–144.

Cragg, B. G., & Ainsworth, A. (1969). The topography of the afferent projections in the circumstriate visual cortex of the monkey studied by the Nauta method. *Vision Research, 9,* 733–747.

Damasio, A. R., Damasio, H., & Van Hoesen, G.W. (1982). Prosopagnosia: Anatomical basis and behavioral mechanisms. *Neurology, 32,* 331–341.

Dean, P. (1976). Effects of inferotemporal lesions on the behavior of monkeys. *Psychological Bulletin, 83,* 41–71.

Dean, P. (1982). Visual behavior in monkeys with inferotemporal lesions. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 587–628). Cambridge, MA: MIT Press.

De Yoe, F. A., & Van Essen, D. C. (1985). Segregation of efferent connections and receptive-field properties in visual area V2 of the Macaque. *Nature, 317,* 58–61.

De Yoe, E. A., & Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neuroscience, 11,* 219–226.

Desimone, R., Albright, T. D., Gross C. G., & Bruce, C. (1984). Stimulus selective properties of inferior temporal cortex in the Macaque. *Journal of Neuroscience, 4,* 2051–2062.

Desimone, R., Fleming, J., & Gross, C. G. (1978). Cortical afferents to inferior temporal cortex in the Macaque. *Assessment Research of Visual Ophthalmology Abstracts, 292.*

Desimone, R., Fleming, J., & Gross, C. G. (1980). Prestriate afferents to inferior temporal cortex: An HRP study. *Brain Research, 184,* 41–55.

Desimone, R., & Gross, C. G. (1979). Visual areas in the temporal cortex of the Macaque. *Brain Research, 178,* 366–380.

Desimone, R., & Ungerleider, L. G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the Macaque. *Journal of Comparative Neurology, 248,* 164–189.

Evarts, E. V., & Tanji, J. (1974). Gating of motor cortex reflexes by prior instruction. *Brain Research, 71,* 479–494.

Feldman, M. F. (1984). Morphology of the neocortical pyramidal neuron. In A. Peters and E. G. Jones (Eds.), *Cerebral cortex: cellular components of the cerebral cortex* (Vol. 1). New York: Plenum Press.

Finkel, L. H., & Edelman, G. M. (1989). Integration of distributed cortical systems by reentry: A computer simulation of interactive functionally segregated visual areas. *Journal of Neuroscience, 9,* 3188–3208.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition un-

affected by shift in position. *Biological Cybernetics, 36,* 193–202.

Fukushima, K. (1988). Neocognitron: A hierarchical network capable of visual pattern recognition. *Neural Networks, 1,* 119–130.

Fukushima, K., & Miyake, S. (1982). Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognition, 15,* 455–469.

Fukushima, K., Miyake, S., & Ito, T. (1983). Neocognitron: A neural network model for a mechanism of visual pattern recognition. *IEEE Transactions on Systems, Man and Cybernetics SMC-13,* 826–834.

Gatass, R., & Gross, C. G. (1981). Visual topography of V2 in the Macaque. *Journal of Comparative Neurology, 201,* 519–540.

Gatass, R., Rosa, M. G. P., Sousa, A. P. B., Piñon, M. C. G. P., Fiorani, M., Neuenschwander, J. S. T., Moura, M. M., Abrahão, J. C. H., & Saraiva, P. E. S. (1990). Visual topography and modular organization of cortical areas in primates. In E. Iwai & M. Mishkin (Eds.), *Vision, memory and the temporal lobe* (pp. 357–367. New York: Elsevier Science Publishing Co.

Gatass, R., Sousa, A. P. B., & Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the Macaque. *Journal of Neuroscience, 8,* 1831–1845.

Gatass, R., Sousa, A. B., & Rosa, M. G. P. (1987). Visual Topography of V1 in Cebus monkey. *Journal of Comparative Neurology, 259,* 529–548.

Gilbert, C. D., & Wiesel, T. N. (1981). Laminar specialization and intracortical connections in cat primary visual cortex. In F. Schmidt, F. Worden, G. Adleman, & S. Dennis (Eds.), *The organization of the cerebral cortex* (pp. 163–191). Cambridge, MA: MIT Press.

Glünder, H. (1987). Invariant description of pictorial patterns via generalized auto-correlation functions. In D. Meyer-Ebrecht (Ed.), *ASST '87, 6. Aachener Symposium für Signaltheorie* (pp. 84–87). Berlin: Springer.

Glünder, H. (1990). Equivalent TLU- and $\Sigma\Pi$-networks for invariant pattern recognition. In *International Neural Network Conference,* July 9–13, Paris.

Gnadt, J. W., & Andersen, R. A. (1986). Spatial, memory, and motor planning properties of saccade-related activity in the lateral intra-parietal area (LIP) of Macaque. *Neuroscience Abstracts, 13,* 454.

Gnadt, J. W., & Andersen, R. A. (1988). Memory related motor planning activity in posterior parietal cortex of Macaque. *Experimental Brain Research, 70,* 216–220.

Gross, C. G. (1973a). Visual functions of inferotemporal cortex. In R. Jung (Ed.), *Handbook of sensory physiology* (Vol. 7, Part 3). Berlin: Springer-Verlag.

Gross, C. G. (1973b). Inferotemporal cortex and vision. In E. Stellar & J. M. Sprague (Eds.), *Progress in physiological psychology* (Vol. 5, pp. 77–123). New York: Academic Press.

Gross, C. G., Bender, D. B., & Mishkin, M. (1977). Contributions of the corpus callosum and the anterior commissure to visual activation of inferior temporal neurons. *Brain Research, 131,* 227–239.

Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science, 166,* 1303–1306.

Gross, C. G., Bruce, C. J., Desimone, R., Fleming, J., & Gatass, R. (1981). Cortical visual areas of the temporal lobe. In C. N. Woolsey (Ed.), *Cortical sensory organization* (Vol. 2, pp. 187–216). Clifton, NJ: Humana.

Gross, C. G., & Mishkin, M. (1977). The neural basis of stimulus equivalence across retinal translation. In S. Harned, R. Doty, J. Jaynes, L. Goldberg, & G. Krauthamer (Eds.), *La-*

teralization in the nervous system (pp. 109–122). New York: Academic Press.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D.B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology, 35,* 96–111.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160,* 106–154.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *Journal of Neurophysiology, 28,* 229–289.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195,* 215–243.

Hubel, D. H., & Wiesel, T. N. (1977). Functional architecture of Macaque monkey visual cortex. *Proceedings of the Royal Society of London, Series B, 198,* 1–59.

Hyvärinen, J. (1981). Regional distribution of functions in parietal association area 7 of the monkey. *Brain Research, 206,* 287–303.

Hyvärinen, J., & Poranen, A. (1974). Function of the parietal associative area 7 as revealed from cellular discharges in alert monkeys. *Brain, 97,* 673–692.

Iwai, E., & Mishkin, M. (1969). Further evidence of the locus of the visual area in the temporal lobe of the monkey. *Experimental Neurology, 25,* 585–594.

Jones, E. G. (1981). Anatomy of cerebral cortex: columnar input-output organization. In F. Schmitt, F. Worden, G. Adleman, & S. Dennis (Eds.), *The organization of the cerebral cortex* (pp. 199–235). Cambridge, MA: MIT Press.

Joynt, R. J., Honch, G. W., Rubi, A. J., & Trudell, R. G. (1985). Occipital lobe syndromes. *Handbook of Clinical Neurology, 1,* 45–62.

Kuypers, H. G., Szwargbart, M. K., Mishkin, M., & Rosvold, H. E. (1965). Occipitotemporal corticocortical connections in the rhesus monkey. *Experimental Neurology, 11,* 245–262.

Lamotte, R. H., & Acuña, C. (1978). Defects in accuracy of reaching after removal of posterior parietal cortex in monkeys. *Brain Research, 139,* 309–326.

Lévy-Schoen, A. (1977). Form identification in peripheral vision. *Optica Acta, 24,* 139–145.

Lynch, J. C., Mountcastle, V. B., Talbot, W. H., & Yin, T. C. T. (1977). Parietal lobe mechanisms for directed visual attention. *Journal of Neurophysiology, 40,* 362–389.

Maunsell, J. H. R., & Van Essen, D. C. (1983). The connections of middle temporal visual area (MT) and their relationship to a cortical hierarchy in the Macaque monkey. *Journal of Neuroscience, 3,* 2563–2586.

Maunsell, J. H. R., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience, 10,* 363–401.

Meadows, J. C. (1974a). Disturbed perception of colors associated with localized cerebral lesions. *Brai, 97,* 615–632.

Meadows, J. C. (1974b). The anatomical basis of prosopagnosia. *Journal of Neurology, Neurosurgery, and Psychiatry, 37,* 489–501.

Mishkin, M. (1954). Visual discrimination performances following partial ablations of the temporal lobe: II. Ventral surface vs. hippocampus. *Journal of Comparative Physiology and Psychology, 47,* 187–193.

Mishkin, M. (1966). Visual mechanisms beyond the striate cortex. In R. Russell (Ed.), *Frontiers of Physiological Psychology* (pp. 93–119). New York: Academic Press.

Mishkin, M. (1972). Cortical visual areas and their interaction. In A. G. Karczmar & J. C. Eccles (Eds.), *The brain and human behavior* (pp. 187–208). Berlin: Springer-Verlag.

Mishkin, M. (1982). A memory system in the monkey. *Philosophical Transactions of the Royal Society of London, Series B, 298,* 85–95.

Mishkin, M. (1990). Vision, memory, and the temporal lobe: Summary and perspective. In E. Iwai and M. Mishkin (Eds.), *Vision, memory and the temporal lobe* (pp. 427–436). New York: Elsevier Science Publishing Co.

Mishkin, M., & Pribam, K. H. (1954). Visual discrimination performance following partial ablations of the temporal lobe: I. Ventral vs. lateral. *Journal of Comparative Physiology and Psychology, 47,* 14–20.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neuroscience, 6,* 414–417.

Miyashita, Y. (1990). Associative representation of the visual long-term memory in the neurons of the primate temporal cortex. In E. Iwai & M. Mishkin (Eds.), *Vision, memory and the temporal lobe* (pp. 75–87). New York: Elsevier Science Publishing Co.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229,* 782–784.

Morel, A., & Bullier, J. (1990). Anatomical segregation of two cortical pathways in the Macaque monkey. *Visual Neuroscience, 4,* 555–578.

Motter, B. C., & Mountcastle, V. B. (1981). The functional properties of the light-sensitive neurons of the posterior parietal cortex studied in waking monkeys: Foveal sparing and opponent vector organization. *Journal of Neuroscience, 1,* 3–26.

Motter, B. C., Steinmetz, M. A., Dufy, C. J., & Mountcastle, V. B. (1987). Functional properties of parietal visual neurons: Mechanisms of directionality along a single axis. *Journal of Neuroscience, 7,* 154–176.

Mountcastle, V. B. (1978). An organizing principle for cerebral function. The unit module and the distributed system. In G. M. Edelman (Ed.), *The mindful brain.* Cambridge, MA: MIT Press.

Mountcastle, V. B., Lynch, J. C., Georgopoulos, A., Sakata, H., & Acuna, C. (1975). Posterior parietal association cortex of the monkey: Command functions for operations within extrapersonal space. *Journal of Neurophysiology, 38,* 871–908.

Neuenschwander, S., Gatass, R., Souza, A. P. B., & Piñon, M. C. G. P. (1990). The PO complex in the Cebus monkey: Location and visuotopic organization. *ECVP, 13th European Conference on Visual Perception,* Paris, p. 378.

O'Regan, J. K., & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research, 23,* 765–768.

O'Regan, J. K., & Nazir, T. (1990). Does human vision have translation invariance. *Spatial Vision, 5,* 81–100.

Pearlman, A. L., Birch, J., & Meadows, J. C. (1979). Cerebral color blindness: An acquired defect in hue discrimination. *Annual Neurology, 5,* 253–261.

Perret, D. I., Mistlin, A. J., & Chitty, A. J. (1987). Visual neurons responsive to faces. *Trends in Neuroscience, 10,* 358–364.

Perret, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research, 47,* 329–342.

Pohl, W. (1973). Dissociation of spatial discrimination deficits following frontal and parietal lesions in monkeys. *Journal of Comparative Physiology and Psychology, 82,* 227–239.

Ratcliff, G., & Davies-Jones, G. A. B. (1972). Defective visual localization in focal brain wounds. *Brain, 95,* 49–60.

Rocha-Miranda, C. E., Bender, D. B., Gross, C. G., & Mishkin, M. (1975). Visual activation of neurons in inferotemporal

cortex depends on striate cortex and forebrain commissures. *Journal of Neurophysiology, 38,* 475–491.

Rolls, E. T., Judge, S. J., & Sanghera, M. K. (1977). Activity of neurons in the inferotemporal cortex of the alert monkey. *Brain Research, 130,* 229–238.

Rosa, M. G. P., Sousa, A. P. B., & Gatass, R. (1988). Representation of the visual field in the second visual area in the Cebus monkey. *Journal of Comparative Neurology, 275,* 326–345.

Rueckl, J. G., Cave, K. R., & Kosslyn, S. M. (1989). Why are "What" and "Where" processed by separate cortical visual systems? A computational investigation. *Journal of Cognitive Neuroscience, 1,* 171–186.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature (London), 323,* 533–536.

Sakata, H., Shibutani, H., Ito, Y., & Tsurugai, K. (1986). Parietal cortical neurons responding to rotary movement of visual stimulus in space. *Experimental Brain Research, 61,* 658–663.

Sakata, H., Shibutani, H., & Kawano, K. (1983). Functional properties of visual tracking neurons in posterior parietal association cortex of the monkey. *Journal of Neurophysiology, 49,* 1364–1380.

Sakata, H., Shibutani, H., Kawano, K., & Harrington, T. L. (1985). Neural mechanisms of space vision in the parietal association cortex of the monkey. *Vision Research, 25,* 453–463.

Sato, T., Kawamura, T., & Iwai, E. (1980). Responsiveness of inferotemporal single units to visual pattern stimuli in monkeys performing discrimination. *Experimental Brain Research, 38,* 313–319.

Seacord, L., Gross, C. G., & Mishkin, M. (1979). Role of inferior temporal cortex in interhemispheric transfer. *Brain Research, 67,* 259–272.

Selfridge, O. G. (1958). Pandemonium: A paradigm for learning. In *Mechanisation of thought processes,* HMSO.

Selfridge, O. G., & Neisser, V. (1963). Pattern recognition by machine. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought* (pp. 406–450). New York: McGraw Hill.

Seltzer, B., & Pandya, D. N. (1980). Converging visual and somatic sensory cortical input to the intraparietal sulcus of the rhesus monkey. *Brain Research, 192,* 339–351.

Shipp, S., & Zeki, S. (1985). Segregation of pathways leading from area V2 to area V4 and V5 of Macaque monkey visual cortex. *Nature (London), 315,* 322–326.

Shipp, S., & Zeki, S. (1989). The organization of connections between area V5 and V2 in Macaque monkey visual cortex. *European Journal of Neuroscience, 1,* 334–354.

Shiwa, T. (1987). Cortico-cortical projections to the monkey temporarl lobe with particular reference to the visual processing pathways. *Archives Italiennes de Biologie, 125,* 139–154.

Steinmetz, M. A., Motter, B. C., Dufy, C. J., & Mountcastle, V. B. (1987). Functional properties of parietal visual neurons: Ra-

dial organization of directionalities within the visual field. *Journal of Neuroscience, 7,* 177–191.

Swindale, N. V. (1990). Is the cerebral cortex modular? *Trends in Neuroscience, 13,* 487–492.

Szentagothai, J. (1975). The "module concept" in cerebral cortex architecture. *Brain Research, 95,* 475–496.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1990). Integration of form, texture and color information in the inferotemporal cortex in the Macaque. In E. Iwai & M. Mishkin (Eds.), *Vision, memory and the temporal lobe* (pp. 101–110). New York: Elsevier Science Publishing Co.

Thorpe, S. (1989). Local vs. distributed coding. *Intellectica,* 1989/2, *8,* 3–40.

Thorpe, S., & Imbert, M. (1989). Biological constraints on connectionist models. In R. Pfeifer, Z. Schreter, F. Fogelman-Soulié, L. Steels (Eds.), *Connectionism in perspective* (pp. 63–92). Amsterdam: Elsevier.

Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychologie Forschung, 31,* 299–337.

Ungerleider, L. G., & Desimone, R. (1986). Cortical connections of visual area MT of the Macaque. *Journal of Comparative Neurology, 248,* 190–220.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior.* Cambridge, MA: MIT Press.

Van Essen, D. C. (1985). Functional organization of primate visual cortex. In A. Peters & F. G. Jones (Eds.), *Cerebral Cortex* (Vol. 3, pp. 359–329). New York: Plenum Press.

Van Essen, D. C., & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neuroscience, 6,* 9.

Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception, 6,* 313–326.

Weiskrantz, L. (1990). Visual prototypes, memory and the inferotemporal cortex. In E. Iwai & M. Mishkin (Eds.),*Vision, memory and the temporal lobe* (pp. 13–28). New York: Elsevier Science Publishing Co.

Weiskrantz, L., & Saunders, R. C. (1984). Impairments of visual object transforms in monkeys, *Brain, 107,* 1033–1072.

Yaginuma, S. (1990). Functional subdivisions of area TE of the inferotemporal cortex in the monkey. In E. Iwai and M. Mishkin (Eds.), *Vision, memory and the temporal lobe* (pp. 29–42). New York: Elsevier Science Publishing Co.

Zeki, S. M. (1969). Representation of central visual fields in prestriate cortex of monkey. *Brain Research, 14,* 271–291.

Zeki, S. M. (1973). Color coding in rhesus monkey prestriate cortex. *Brain Research, 53,* 422–427.

Zeki, S. M., & Shipp, S. (1988). The functional logic of cortical connections. *Nature (London), 335,* 311–316.

Zihl, J., Von Cramon, D., & Mai, N. (1983). Selective disturbance of movement vision after bilateral brain damage. *Brain, 106,* 313–340.