# Chapter 12
# Self-Directed Learning in a Complex Environment

It has been conjectured that a template-filter model for cognitive processes such as pattern learning and recognition would not be able to operate effectively in a complex visual environment because nearby or partially occluding objects would cause insurmountable interference with the processing of target objects (Pinker 1984). Early tests of this conjecture suggested that the limitations attributed to mechanisms of this kind were not as severe as they were represented to be (Trehub 1986). Additional evidence has been provided in simulations of the brain model, where the operation of adaptive filter cells (comb filters) in the detection matrix is critical. These simulations indicate that the hypothesized parsing and filtering mechanisms for pattern learning and recognition can perform effectively in a complex environment. The results also demonstrate that when a novelty detection circuit modulates the excitatory bias of a synaptic matrix, the model is able, in a completely self-directed fashion, to learn, recall, image, and find the objects that compose its fortuitous visual environments.

## *Parsing and Learning*

The simulated brain system consisted of a 16 × 16–cell retina, the neuronal modules for scene representation and visual parsing described in chapter 7, a basic synaptic matrix for pattern detection and imaging, and a novelty detector. The contribution of axon transfer factor (ATF) to synaptic modification was arbitrarily set at $c = 1.0$. The contribution of dendrite transfer factor (DTF) was set at $k = 100$. The gradient coefficient of dendrodendritic excitation of neighboring cells in the mosaic array was set at 0.6.

Outdoor (far) and indoor (near) visual "environments" were created in sketch-to-pixel conversions. These complex scenes were then presented to the simulated neuronal system for parsing, learning, and object recognition.

*Simulation 1*
The model was exposed to a succession of different outdoor environments, each consisting of a variety of trees, houses, buildings, cars, and animals and some including the outline of distant hills (figure 12.1). Because the actual coordinate projections from a 3-D retinoid to the mosaic cells of a synaptic matrix maintain size constancy as viewing distance changes, the distance of an object as well as its retinal image will determine the pattern of synaptic transfer weights ($\phi$) on any filter cell that is modified at the time of learning. Since the simulation did not incorporate mechanisms of stereopsis or visual accommodation, the viewing distance to the major elements of each scene was assumed to be constant.

Environmental scenes were presented in a frame of 4000 pixels. Thus, there would be 4000 potential fixation points during the viewing of each scene if there were no principled constraints on fixation in an extended visual environment. Given the proposed brain model, however, the number of potential fixations on any complex scene is greatly reduced by contour flux control of saccadic action and by the centroid-finding properties of the retinoid system. A significant consequence of this is to reduce the number of object exemplars that must be learned for adequate recognition performance in new visual environments.

In the previous simulations, visual learning occurred only when a teacher (the operator) indicated that an incorrect recognition response was made by the detection matrix. In this simulation, visual learning was initiated solely by the model brain mechanism whenever it "judged" an object (or a part of an object) to be novel. If the latency of discharge for the first class cell ($\Omega$) in the detection matrix to respond to a simulus exceeded the delay time of the novelty detector, then the synaptic matrix received a pulse of excitatory bias and the current exemplar was automatically learned. Feedback from a teacher or from an error-detecting module was not required, nor was it given.

Another aspect of the model's performance that should be noted is the manner in which it associated common names to the object tokens that it had learned. Periodically, after the exploration of each new environment, the system drew an image (on the CRT screen) of each object (or part of an object) that it had learned to represent by a biological "name" (a class cell token) but for which it had no common name. It then asked what that object is called. If a name was provided, the neuronal network automatically entered that name (character string) in its lexicon and assigned it to the class cell token of its referent.

At the start of each scene-parsing operation, the model first fixated

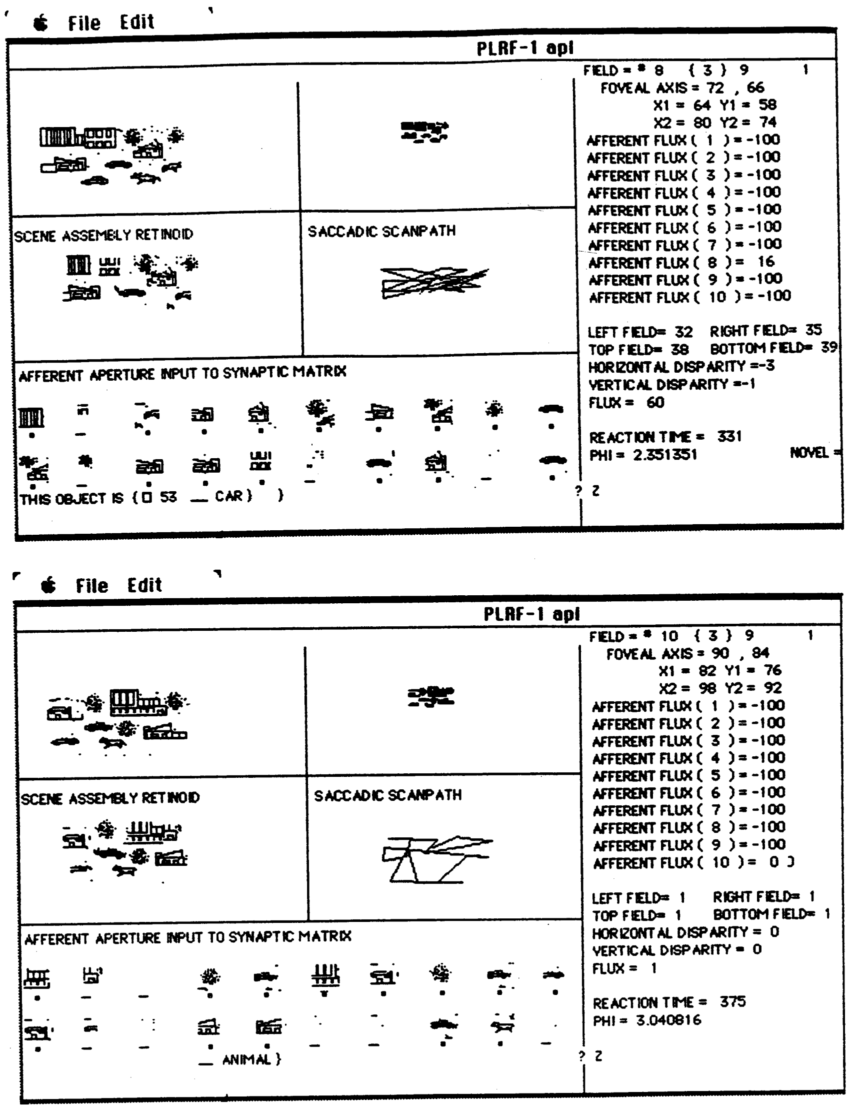



Figure 12.1
Visual exploration of two outdoor scenes (printout of CRT displays). Upper left panel in each display shows scene presented. Central panel shows scanpath of 20 saccades. Middle left panel shows reconstruction on scene assembly retinoid. Bottom panel shows images parsed after each saccade. Dots mark images that were captured and gated to synaptic matrix to be recognized and (if novel) learned. Short horizontal lines mark parsed images that were not captured.

on the retinotopic locus of the flux detector with maximum output. The afferent field aperture then closed to its fully constricted state, arbitrarily set at six retinal units in width and height. The fully expanded afferent aperture was limited to $16 \times 16$ retinal units. Starting error tolerance for quadrantal disparity over either horizontal or vertical axes was set at three units. At any fixed aperture, if error tolerance was exceeded on a given axis, the pattern of excitation on the retinoid was automatically shifted in the appropriate direction to reduce hemifield disparity on that axis. When pattern position satisfied error tolerance for one axis, the pattern was automatically shifted on the other axis unless it was already within tolerance. If shifting the image on the second axis now resulted in an unacceptable error on the first, error tolerance was automatically relaxed one unit. Whenever the retinoid pattern was brought within tolerance for both horizontal and vertical disparities, the afferent aperture expanded one unit, and the process repeated automatically until full aperture was achieved. This operation was assumed to involve an expenditure of processing effort; if a pattern translation of nine units on any retinoid axis did not bring disparity within tolerance, the system stopped trying at its current fixation and initiated a saccade to the next highest flux region. Whenever the afferent aperture reached the state of full expansion with an input flux of at least 25 active cells and with disparity within tolerance on both axes (thus centered on the normal foveal axis), the excitation pattern on the retinoid was gated to the synaptic matrix for recognition. Thus, each excitation pattern projected to the synaptic matrix was a discrete image parsed out of the complex visual field.

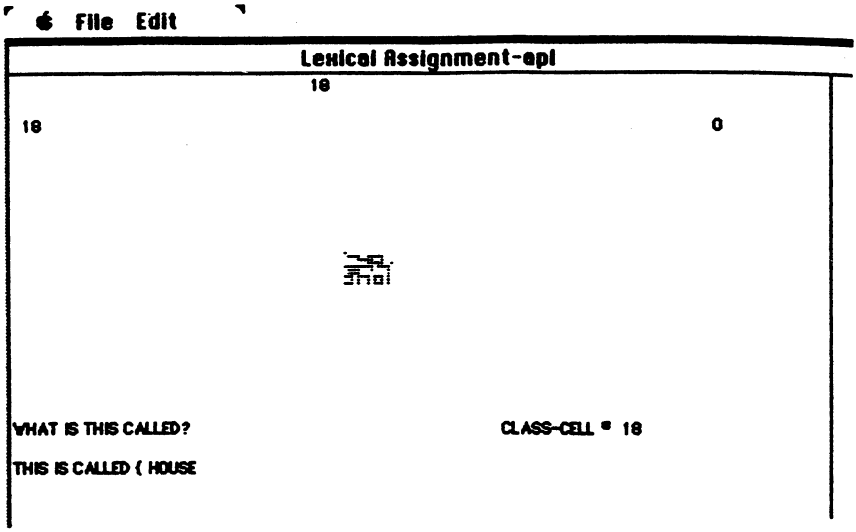
The simulated visual system was allowed to make 20 saccades during the presentation of each scene. Because of the requirement for quadrantal balance before gating from the retinoid module to the synaptic matrix, an image that falls on the line of sight at the terminus of a saccade is rarely the image projected to the mosaic cell array of the synaptic matrix. Here, it will be useful to make a distinction between visual fixation, which refers to any resting locus of gaze in the visual field, and visual capture, which refers to the bringing of an object's visual centroid to the retinoid's normal foveal axis, within tolerance for error and effort. Only images that have been captured are gated to the synaptic matrix (where they can be learned if filter cells and mosaic cells receive sufficient excitatory priming).

Threshold for stimulus novelty varied from moment to moment in a random fashion and corresponded to a recognition latency (first-order class cell [$\Omega$] response) ranging from 330 to 360 milliseconds. If the recognition response time exceeded the current novelty threshold, the synaptic matrix was primed, and the captured exemplar was learned.

Through learning during the course of the simulation, the model visual system was building a repertoire of class cell tokens (labeled lines in the detection matrix) and latent images (class cell collaterals on mosaic cells in the imaging matrix), which represented the objects or parts of objects parsed and captured in each environment. In order to facilitate communication by providing a common name as well as a biological name for the things it had learned, the model performed as follows. After completing the exploration of each scene, it traced from its imaging matrix one image at a time on the CRT screen. The images traced were those exemplars that it had just learned but for which it had no common name. After drawing an image, it asked what that object is called (figure 12.2). When the name (character string) was provided, it was linked to the currently active class cell, the neuronal token of the object in question. As the simulation progressed, the model was able to signal recognition of an increasing number of objects by their common names as well as by their neuronal tokens.

The simulation was terminated after 160 exemplars (objects or parts of objects) were captured. At this point, 12 different outdoor scenes had been presented, and, of the 160 exemplars that had been captured, 82 had exceeded the novelty threshold and had been learned. The examples given in figure 12.1 are two printouts of the CRT screen during the simulation. The upper left panel in each printout shows the actual scene presented to the model system for visual parsing, recognition, and learning. The central panel shows the scanpath of the 20 saccades that were initiated in the course of visual exploration. The middle left panel shows the partial neuronal reconstruction of the full visual field on a scene assembly retinoid. This fragmentary visual representation was created by the reassembly of the excitation patterns (images) that the model had parsed during the course of 20 successive saccades over the current scene. In the large bottom panel are the successive images that were parsed after each saccade. Only images marked by a dot beneath them were actually captured and transmitted as input to the synaptic matrix for recognition and, if novel, for learning. Images marked by a short horizontal line beneath them were those retinal stimuli that the system was unable to capture within the constraints of centroid tolerance or parsing effort.

Of 20 saccadic fixations over each of the outdoor scenes, the number of images captured ranged from 9 (45 percent) to 17 (85 percent), with an overall capture rate of 67 percent. Because 82 of these stimuli had exceeded the threshold of novelty, a corresponding number of



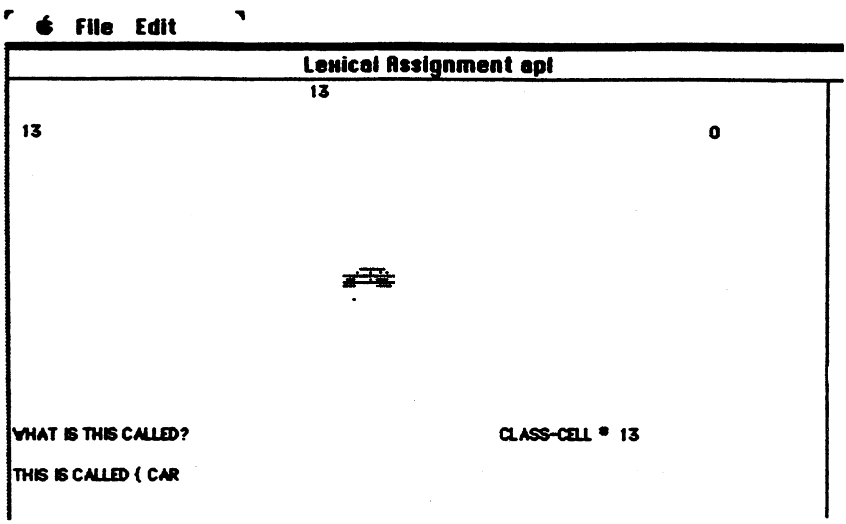




Figure 12.2
Two instances of lexical assignment. Model displays the image evoked by discharge of each class cell token for which it has not yet learned a common name and asks what the image is called. Operator provides the appropriate name (bottom left).

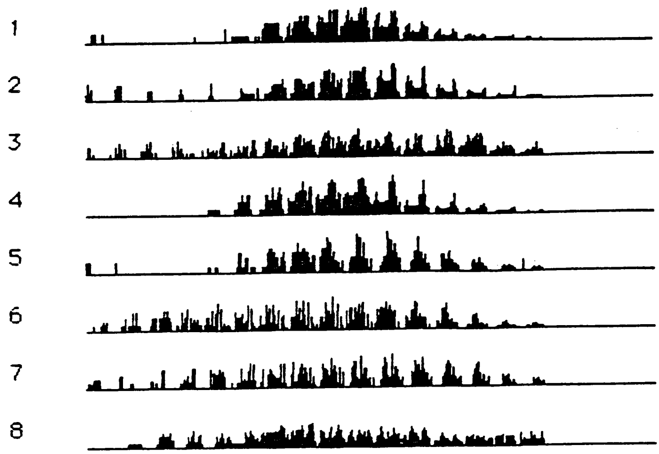



Figure 12.3
Examples of the distribution of synaptic transfer weights (ϕ) on dendrites of filter cells that have learned images captured during exploration of outdoor scenes. Each point on the dendritic line represents a particular synaptic location. Amplitude of each vertical line represents relative magnitude of ϕ for that synapse. The objects learned by these cells are as follows: (1) a car, (2) an animal, (3) a house, (4) a different exemplar of a car, (5) a different exemplar of an animal, (6) a tree, (7) a different exemplar of a tree, (8) a building.

filter cells in the detection matrix had been synaptically modified in accordance with the learning formula. Figure 12.3 shows the synaptic transfer weight (ϕ) profiles of an arbitrary sample of the 82 learning-modified filter cells. Differences among the ϕ distributions over the population of filter cells in the detection matrix determine the selectivity of object recognition. Associated ϕ distributions in the imaging matrix shape the images that are evoked on the mosaic cell array by active class cell collaterals.

The course of recognition performance was assessed by examining the percentage of correct responses in successive blocks of 20 captured images. Figure 12.4 shows the improvement in recognition as the number of objects parsed and captured increased to the simulation limit of 160. The curve of performance increases rapidly and appears to reach a plateau at a level between 80 percent and 90 percent correct recognitions.

*Simulation 2*

In the second test, the model visual system was exposed to a succession of cluttered desktop scenes, each having different exemplars and arrangements of books, bookmarks, telephones, ashtrays, and

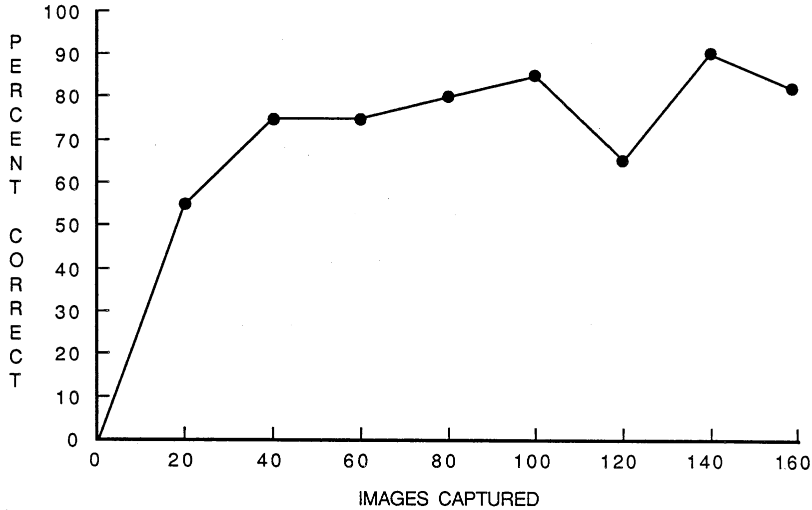



Figure 12.4
Graph showing the percentage of correct object recognition responses in successive blocks of 20, plotted against the cumulative number of images captured in the outdoor scenes.

pencils. The model was required to parse, learn, and recognize the objects that composed its visual environment. Like the previous simulation, learning was automatically initiated by the model whenever its class cell response time exceeded the delay time (threshold) of its novelty detection circuit. Common names were provided in response to the model's queries on the basis of images drawn after the exploration of each new scene. All characteristics of the neuronal mechanisms (number of retinal cells, parameters for parsing and capture, learning) were the same as in the first simulation.

After 13 different desktop scenes had been presented, the model had captured 160 images, and the simulation was terminated. Of the 160 exemplars that had been captured, 61 had exceeded the novelty threshold and had been learned. Figure 12.5 shows two printouts of the CRT screen during this simulation test.

The number of images captured during the 20 saccadic fixations on each of the desktop scenes ranged from 8 (40 percent) to 15 (75 percent). The overall capture rate was 62 percent. Because 61 of these stimuli had exceeded the threshold of novelty, a corresponding number of filter cells in the detection matrix were synaptically modified. Profiles of synaptic transfer weights ($\phi$) on a sample of eight of the 61 learning-modified filter cells are shown in figure 12.6.

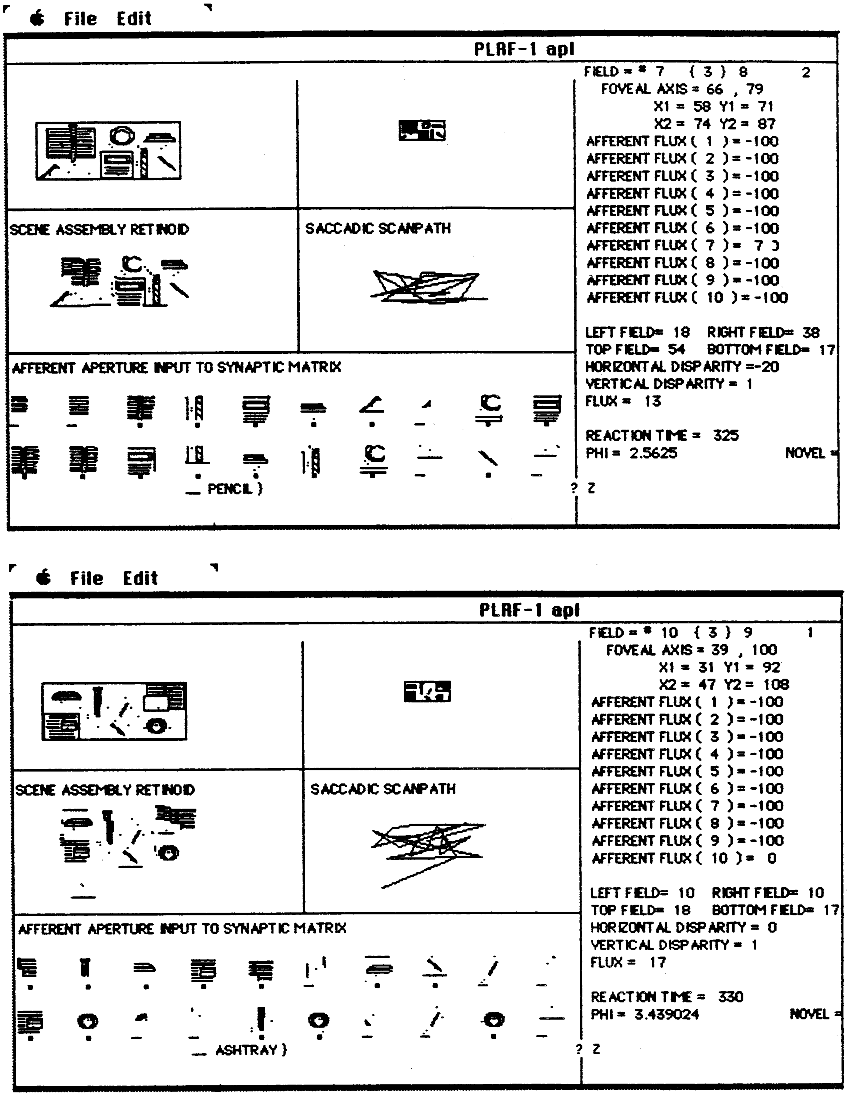



Figure 12.5
Visual exploration of desktop scenes (printout of CRT displays). Upper left panel in each display shows scene presented. Notice that books are partially occluded by a bookmark (upper display) and a sheet of paper (lower display). Central panel shows scanpath of 20 saccades. Middle left panel shows reconstruction on scene assembly retinoid. Bottom panel shows images parsed after each saccade. Dots mark images that were captured and gated to synaptic matrix to be recognized and (if novel) learned. Short horizontal lines mark parsed images that were not captured.

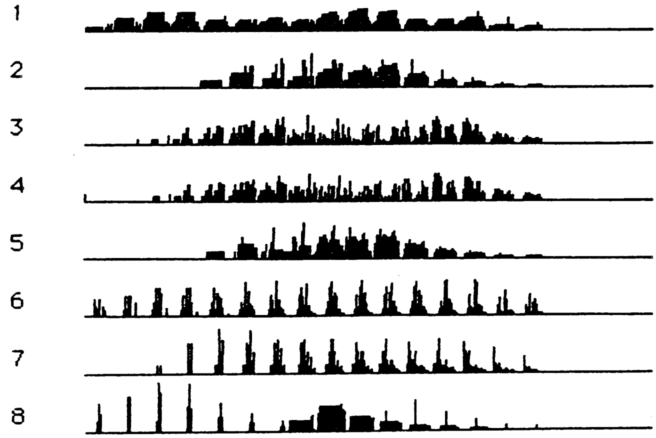



Figure 12.6
Examples of the distribution of synaptic transfer weights ($\phi$) on dendrites of filter cells that have learned images captured during exploration of desktop scenes. The objects learned by these cells are as follows: (1) a book, (2) a telephone, (3) an ashtray, (4) a different exemplar of an ashtray, (5) a different exemplar of a telephone, (6) a bookmark, (7) a pencil, (8) an edge of the desk.

The course of recognition performance was assessed by examining the percentage of correct responses in successive blocks of 20 captured images. Figure 12.7 shows the improvement in recognition as the number of objects parsed and captured increased to the simulation limit of 160. The curve of performance increases rapidly and then tends to flatten, reaching a level of 95 percent correct recognitions.

*Searching for Objects*

The ability of the model visual system to search for and find named objects in complex scenes was also demonstrated. In the simulations of search behavior, two neuronal processes were automatically triggered. First, whenever the model system was in a search mode, all cells in the mosaic array received a sustained increment of inhibition ($-1$ on each cell). Second, when the name (character string) of an object to be found matched a common name linked to class cell tokens in the detection matrix, those class cells that corresponded to the named object received a sustained increment of excitation, causing their axon collaterals to induce a pattern of net positive bias on just those mosaic cells in the imaging matrix that composed their associated images. These neuronal processes have the joint effect of suppressing the response of filter cells that are not tuned to the object

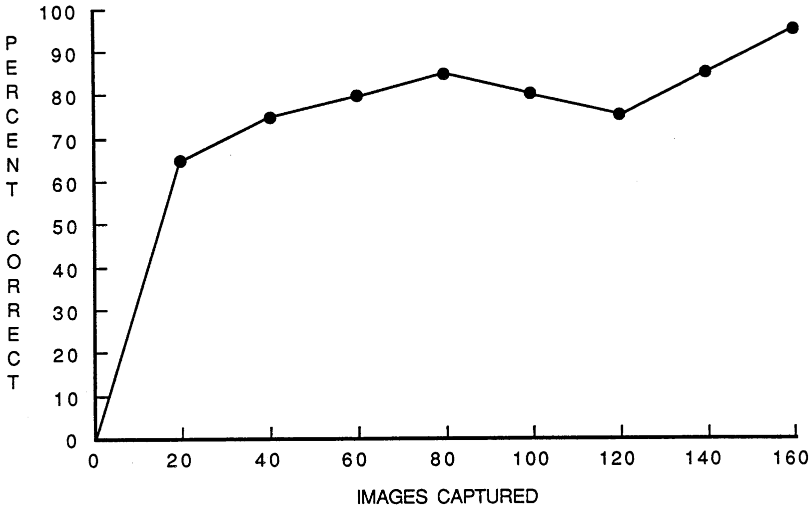



Figure 12.7
Graph showing the percentage of correct object recognition responses in successive blocks of 20, plotted against the cumulative number of images captured in the desktop scenes.

of search and facilitating an appropriate response to the searched-for object.

Saccadic activity, object parsing, and capture proceeded in the same way as in the previous simulations. However, if a captured image was not identified as the object of search, a new saccade was immediately initiated, and search continued. When a captured image was recognized as the object to be found, it was named, and its neuronal representation and relative spatial location were marked on the scene assembly retinoid. Figure 12.8 shows two CRT screen printouts when the model was "asked" to find a car and a tree in outdoor scenes. Figure 12.9 shows the model's responses on the CRT when it was asked to find an ashtray and a telephone in desktop scenes.

### *Comments*

The simulated brain mechanisms parsed objects out of complex environments and learned to recognize them at a level of accuracy between 80 and 95 percent without the assistance of a teacher or feedback from an error-detecting source; reconstructed rough representations of environmental scenes by assembling, at their relative

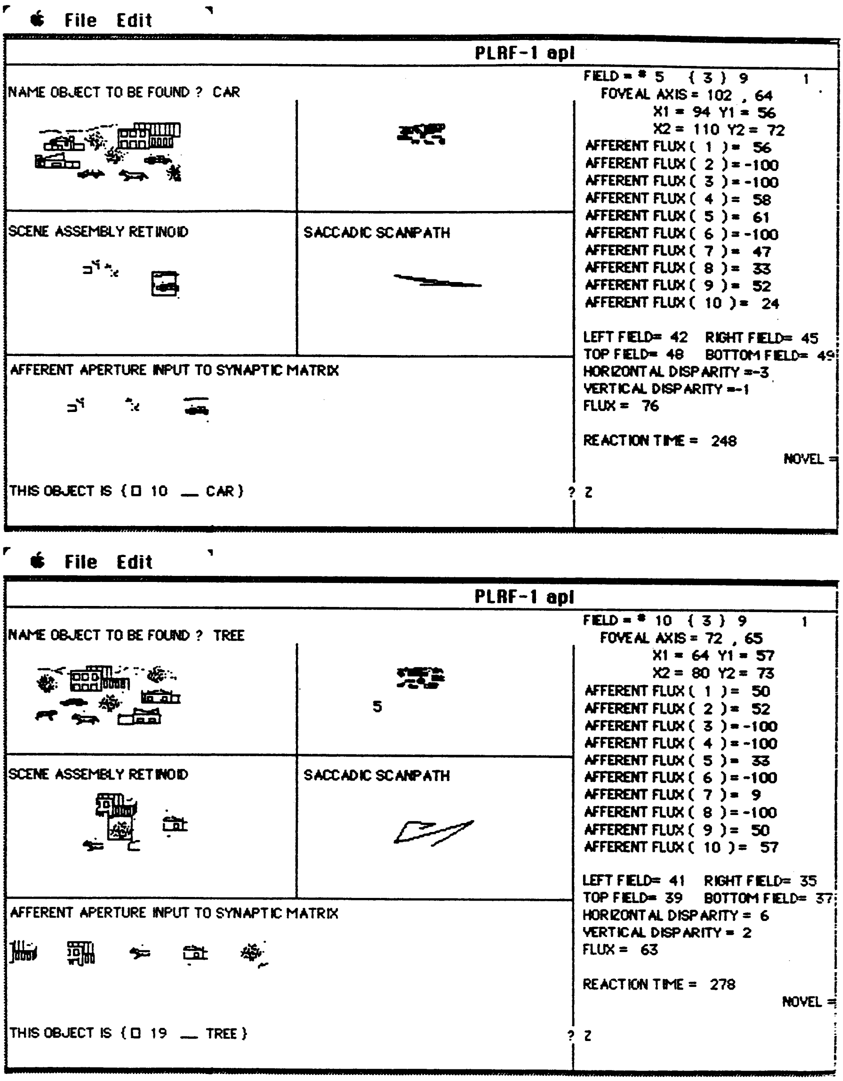
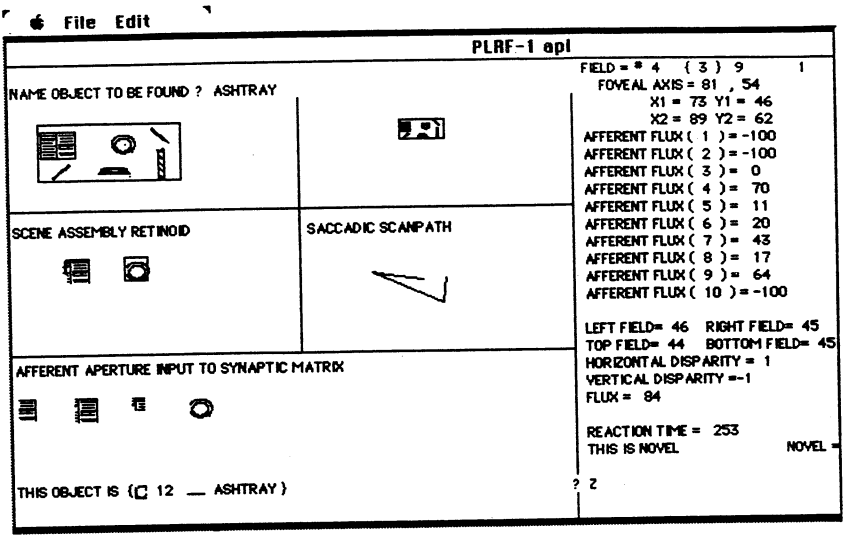



Figure 12.8
Active search. Small rectangular frame around parsed object on the scene assembly retinoid indicates that a searched-for object has been found and its spatial location registered. In these examples, the model was asked to find a car (top display) and a tree (bottom display).



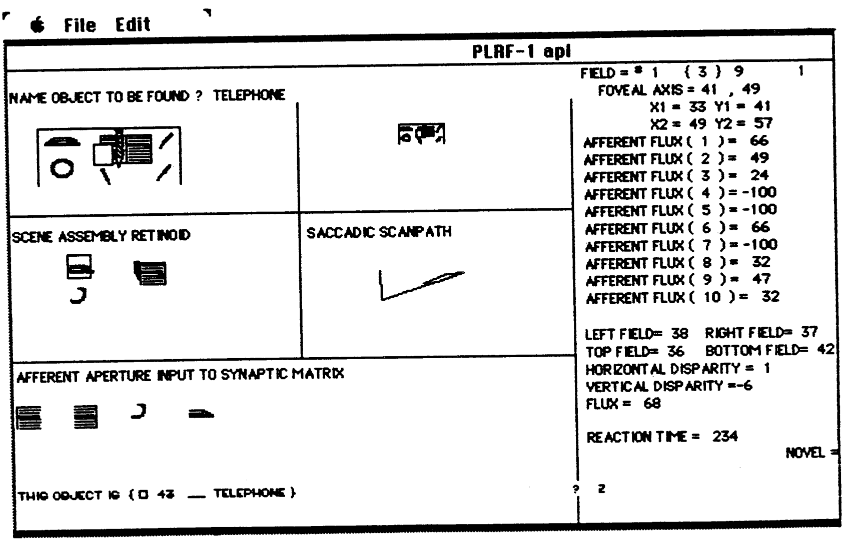




Figure 12.9
Active search. The model finds an ashtray (top display) and a telephone (bottom display).

spatial coordinates on a retinoid, the successive images of objects that it had parsed; and searched for, found, and represented in retinoid space the location of objects in its visual environment when they were designated by name. Given these results, it seems fair to conclude that brain mechanisms of the template-filter kind can effectively serve pattern learning and recognition in complex visual environments.

The processes by which the model assembles a spatially integrated neuronal representation of a frontal scene from a sequence of disparate fixations on the environment might provide an explanation for the clinical findings of hemispatial neglect in some brain-injured patients (Bisiach, Luzzati, and Perani 1979; Bisiach and Berti 1989). Suppose there were selective damage to a hemifield of autaptic cells in the retinoids or neuronal failure for one direction of the shift control mechanism. In the former case, the individual would be unable to experience beyond the retinal level any representation of environmental space homologous with the damaged area of the retinoids. In the latter case, an individual would be unable to translate to the normal foveal axis any image evoked on one of the retinoid hemifields. For example, if the shift-right circuit were damaged, an autaptic image in the left hemifield could not be shifted rightward to the normal foveal axis. Inability to bring an image to the normal foveal axis means that the image cannot be recognized in the detection matrix.

The demonstrated ability of the putative brain system to construct a neuronal analog of the spatial layout of objects in a visual environment supports the theory that candidate plans of action can be instantiated and tested covertly by heuristic excursions of the self-locus in retinoid space. Significant spatial locations, paths to goal objects, relative distances between objects, barriers to direct access, and so forth can be computed and represented analogically by patterns of autaptic cell activity on the retinoid substrate. Such representations can then be learned in synaptic matrices and stored as part of an enduring knowledge base.

In these simulations, the tasks of parsing, learning, and recognition were performed in artificial 2-D visual environments. One might question how well the model would perform in complex 3-D environments. It would clearly be desirable to run simulations in natural settings using photosensitive binocular sensors for retinal input. Such tests would require a computer powerful enough to model effectively the hypothesized 3-D brain mechanisms as integral parts of the simulated visual-cognitive system.