Chapter 7
Analysis and Representation of Object Relations

Among the most challenging problems that must be solved by any theoretical model purporting to explain the competence of the human brain for ecologically relevant tasks are those relating to the analysis and representation of the internal structure in an extended spatial layout of multiple objects. For example:

- How can a complex visual environment be effectively parsed into objects that can be learned and represented as cognitive entities?
- How can we selectively remember the location of objects in 3-D space?
- How can we extract and represent spatial relationships among objects?
- How can we represent the movement of a selected object?

In this chapter, we will consider some plausible brain structures that I believe can provide answers to these questions.

Parsing Objects

Try to imagine a person with an intact visual system but absolutely no knowledge about objects in the visual world. In such a case, since any point of gaze would be no more meaningful than another, where would one look? How could the person parse out the articulated objects and parts of the observed environment (Hoffman and Richards 1984, Pinker 1984)? If an individual were unable to capture significant parts of the world as distinct sensory entities, how could they be learned and added to an evolving knowledge base? It is clear that the ability to parse neuronal representations of objects systematically out of a continuous flux of visual stimulation is critical for the development of cognitive competence. The neuronal system described next addresses this problem; it can parse objects in a complex visual environment (Trehub 1986).

Architecture of a Parsing System
Shown in figure 7.1 is a block-flow diagram of the various modular mechanisms already proposed, with the visual parsing system incorporated in this architecture. The principal processing elements in the model can be outlined as follows:

1. Center-surround mechanisms in the retina and lower-level visual nuclei extract contours from the light-intensity distribution of the visual field.

2. There are cells that independently integrate contour excitation within small, discrete regions over the entire visual field. These flux detectors are retinotopically indexed and serve to drive visual saccades to regions of maximum contour flux.

3. A visual field constriction mechanism can limit the effective stimulus input to an area of variable retinal diameter centered on the foveal axis.

4. A postretinal dynamic visual buffer, called a retinoid, can translate patterns of retinal stimulation over an egocentric coordinate space. This module automatically locates and positions pattern centroids on a standard reference axis (the normal foveal axis) within the visual system.

It is assumed that the total visual field is analyzed by an array of retinotopically indexed cells, each of which receives input from a relatively small region of the complete retinal field (figure 7.2). These are the flux detectors, and each of these cells integrates the amount of visual contour excitation in its particular retinal region and discharges with a frequency proportional to its total excitation. It is assumed that the receptive field of each detector overlaps the receptive fields of its neighboring detectors. These contour flux detectors feed a matched array of cells that control visual saccades. The flux detector with the highest discharge frequency captures control of the saccadic apparatus (by the same kind of mechanism that causes the most active class cell to determine the output of the detection matrix) and directs a saccade to the center of the circumscribed visual region that provides its input. Thus, the region of visual space with the highest contour density will be fixated first.

At the start of the parsing process, the visual field aperture is constricted to a small window on the environmental scene, and an initial tolerance level is set for hemifield mismatch in the retinoid system before the centroid of the current effective visual stimulus pattern is shifted to the normal foveal axis. The visual aperture is progressively enlarged in a stepwise fashion and at the same time relaxes its tolerance for hemifield discrepancies (error due to quadrantal-excitation

ENVIRONMENT

Retinas
and
Preprocessors

Field Constrictor

Retinoid
System

Size Transformer

Rotation
Transformer

Field Constrictor

Synaptic Matrix
(visual pattern recognition)

mosaic array
imaging matrix
detection matrix

Novelty Detector
(arousal)

Synaptic Matrix
(lexical assignment)

mosaic array
imaging matrix
detection matrix

Semantic Network
(lexical association and
proposition-based inference)

word cells
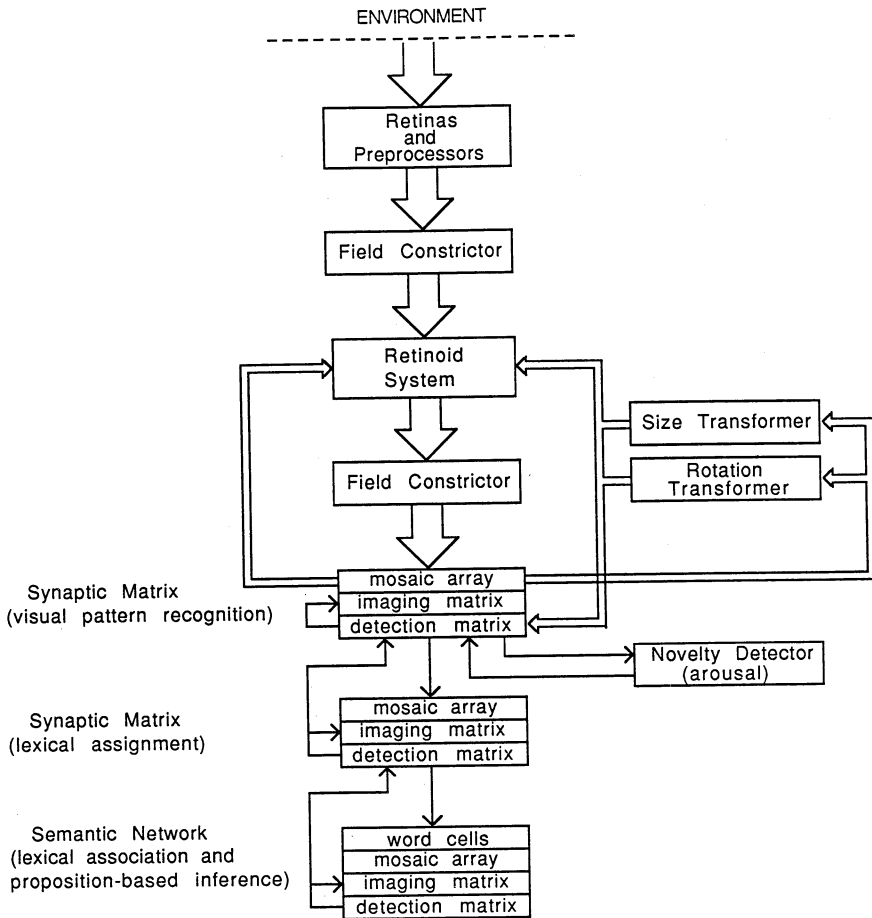mosaic array
imaging matrix
detection matrix

Figure 7.1
Block-flow diagram of visual system and semantic network. Double-lined channels transfer large multicell patterns of excitation. Single-lined channels transfer labeled-line excitation.
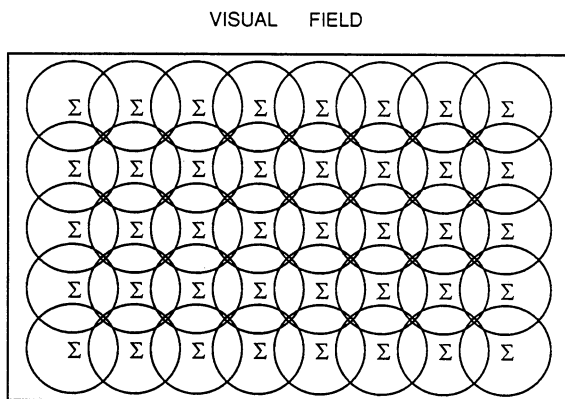
VISUAL   FIELD

Figure 7.2
Contour flux detectors with overlapping summation fields. Contour excitation is independently integrated over the limited visual field of each detector. The flux detector with the highest output drives a self-locus excursion and a saccade to the retinotopic locus of its visual field.

imbalances). At each step, the system seeks to adjust the current centroid of the stimulus component within the larger pattern aperture so that it lies on the normal foveal axis of the retinoid layer that projects to the synaptic matrix. When the afferent aperture reaches a limiting size, the pattern of retinoid excitation on the normal foveal axis is gated to the synaptic matrix for learning or recognition. Then the currently dominant flux detector is inhibited. This allows the detector with the next highest level of activity to direct the next saccade, causing a new fixation, and the parsing process is repeated. The processing sequence for object parsing is illustrated in figure 7.3.

An Example of Performance
The model was tested for its ability to parse objects from a 2-D projection of the visual environment. In the following simulation, it was provided with a 16 × 16–cell retinal fovea and 20 flux detection cells with overlapping receptive fields that combined to cover the entire visual scene. The exact loci of detection cell fields were distributed in quasi-random fashion over the scene. The initial afferent aperture was arbitrarily set at four retinal units in width, and the fully expanded aperture was limited to 16 retinal units. Initial error tolerance for centroid positioning was arbitrarily set at two units of hemifield disparity. The flowchart for the computer simulation is shown in figure 7.4.
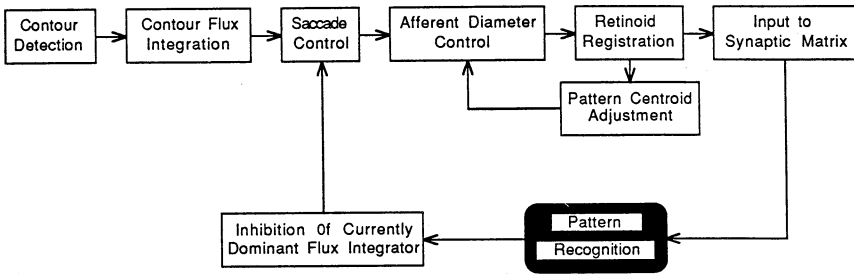
Processing   Sequence

Figure 7.3
Processing sequence for parsing an object in a complex visual environment.

The objects to be parsed were randomly located in a 100×40–pixel visual scene. The stimuli were alphabetical characters, a human face, and a person seated at a table with just the head, upper torso, arms, and hands visible. At the start of each scene-parsing operation, the model first made a saccade to the retinotopic locus of the flux detector showing the highest activity. Then initial error tolerance was set ($+/-$ 2 retinal units of hemifield disparity) and the afferent field aperture closed to its initial constriction state (4 retinal units). The centroid of the retinal pattern that fell within the afferent aperture was then moved to the normal foveal axis by the retinoid mechanism described in chapter 4. At any fixed aperture, if error tolerance was exceeded on a given axis, the retinoid pattern shifted in the appropriate direction to reduce disparity on that axis. When the pattern position satisfied error tolerance for one axis, the pattern shifted in the appropriate direction on the other axis unless it was already within tolerance. If shifting the image on the second axis resulted in an unacceptable error on the first, error tolerance relaxed one unit. Whenever the pattern was brought within axial tolerance for both horizontal and vertical disparities, the afferent aperture expanded one unit and the process repeated until full aperture was achieved. This operation was assumed to involve an expenditure of processing effort, and if a retinoid shift over nine consecutive cells on any axis did not reduce hemifield disparity on that axis, the system stopped trying at its current fixation and initiated a saccade to the next highest flux region.

In one test of the parsing system, the letters *A* and *B* were presented at various locations in the visual field, and the ability of the system to isolate and capture these objects on the normal foveal axis
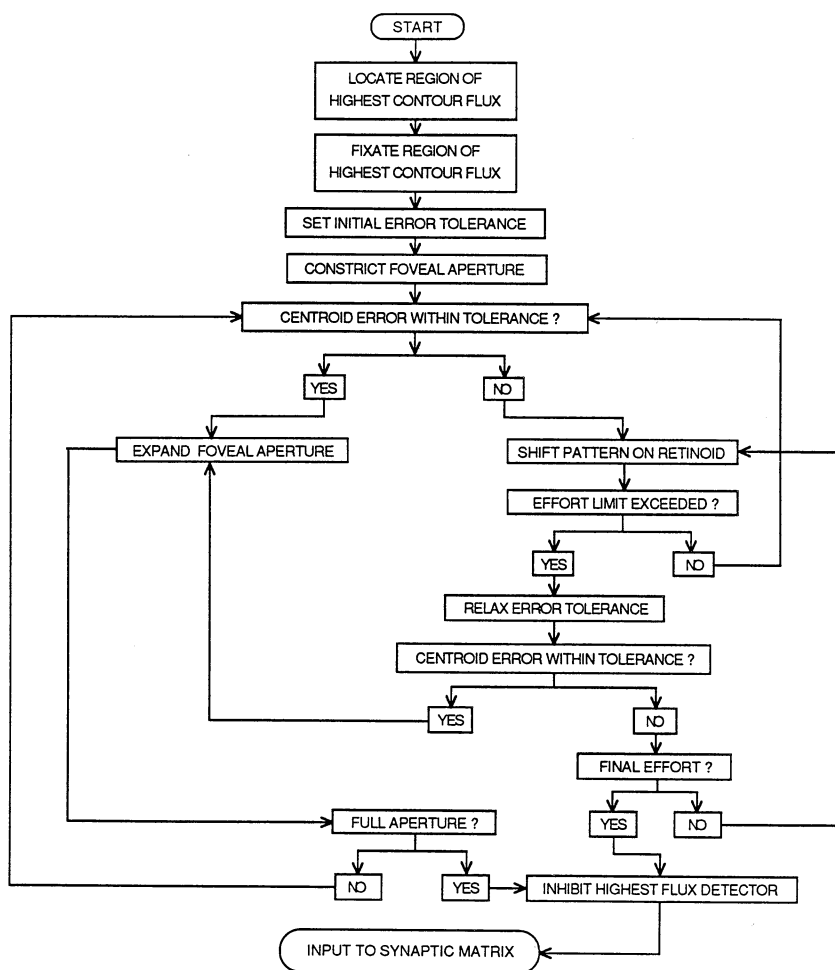
Figure 7.4
Flowchart for object-parsing simulation.

was successfully demonstrated. Another test required the visual system to "look at" and parse displays of a larger number of upper- and lowercase letters. Figures 7.5, 7.6, and 7.7 are printouts taken directly from the CRT. They indicate that even in a visual environment cluttered with many objects, a parsing system of this kind can operate effectively.

These tasks required the visual system to parse individual objects (alphabetic characters) out of a display consisting of a number of randomly located objects, each small enough to be represented in its complete form within the receptive field of the fully expanded afferent aperture. What happens when an object is so large or viewed so near that the visual angle it subtends is much larger than the retinal fovea and the visual projection of its outer bounds extends beyond the afferent aperture? In such cases, we would like the system sys-

Figure 7.5
CRT display of object-parsing simulation. Upper-left panel shows objects A and B as located in the "environment." Middle panel shows the path of successive saccadic fixations as parsing proceeds. Bottom panel shows successive enlargement of afferent aperture and automatic adjustment of object centroid to normal foveal axis following each of eight saccades. Empty or absent apertures indicate that an object centroid could not be adjusted within tolerance on that fixation at that aperture. An empty maximum aperture indicates that no acceptable centroid adjustment was made before the effort limit was exceeded and a new saccade was initiated. Left-middle panel shows a reconstruction of the visual environment on the scene assembly retinoid.
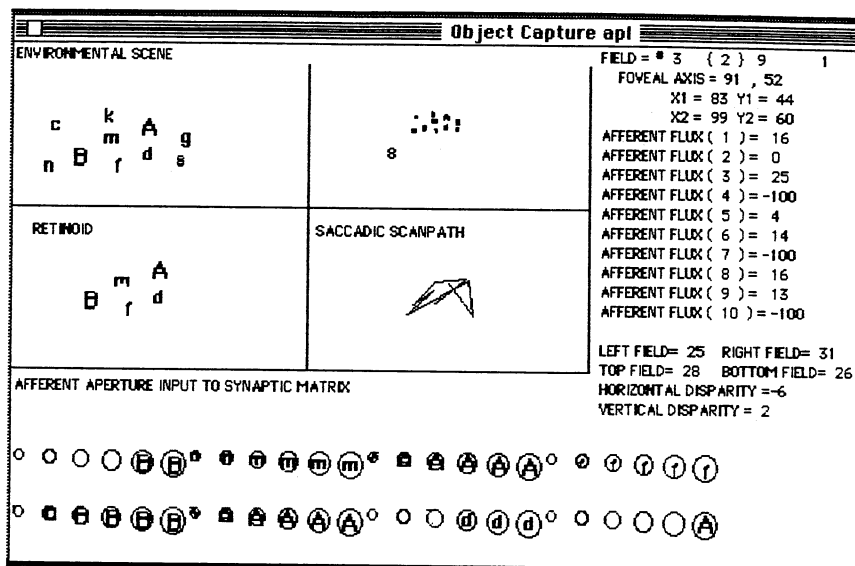
Figure 7.6
CRT display of object-parsing simulation. Upper-left panel shows an "environment" containing 10 objects. Only the five objects that were parsed in the course of eight saccadic fixations are represented on the scene assembly retinoid.

tematically to capture parts of the larger object. To test the behavior of the model in situations of this kind, single stimuli that were much larger than the maximum afferent aperture were presented for parsing—in one case, a human head and face was shown and in another a person seated at a table with the upper part of the body, arms, and hands visible. CRT displays of these simulations are shown in figures 7.8 and 7.9.

When the head and face was presented, the visual system isolated and captured the following parts: left eye (twice); right eye and bridge of the nose (twice); right jaw line, neck, and shoulder; and mouth and lower nose (three times) (figure 7.8). The left-middle panel in figure 7.8 shows the retinotopic excitation pattern that is formed, one part at a time, in the scene assembly retinoid on the basis of the successive saccadic fixations through the afferent aperture. When the figure of a person seated at a table was presented, these parts were isolated: face (twice), hands (three times), edge of table, and right forearm (figure 7.9). Again, notice the image composed on the scene assembly retinoid by successive fixations through the afferent aperture.
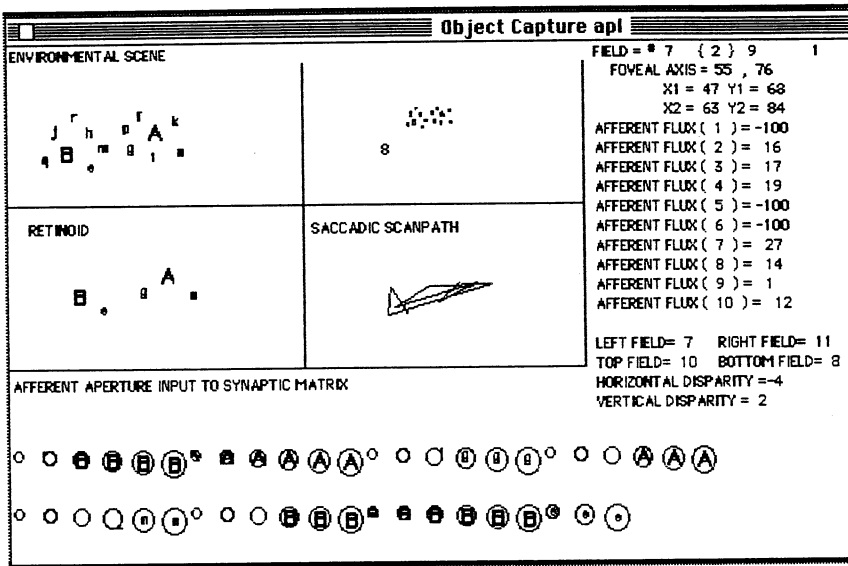
Figure 7.7
CRT display of object-parsing simulation. Upper-left panel shows an environment containing 14 objects. Notice that parsing is effective even in a densely cluttered visual environment.

At each parsing operation, the momentary size of the afferent aperture will limit the size of an object (or part of an object) in the larger visual field that can be projected to the synaptic matrix. For example, if the afferent aperture were limited to a small expansion, an eye or mouth could be isolated as a part of a face to be recognized or learned as a unitary feature. A larger expansion would capture the image of an entire face; this could then be matched as a holistic form against filter cells in the detection matrix or, if novel, provide the excitation pattern for a new profile of synaptic weights in a previously unmodified filter cell.

In this simulation of a minimal parsing mechanism, the relative density of visual contours over segments of the frontal scene served to control successive saccades. It is assumed that in the complete visual-cognitive system, low-level processes such as response to an abrupt change in illumination or to sudden motion and higher-level processes such as a preprogrammed search routine can also select the locus of gaze. But no matter what drives the saccade, the system performs a subsequent fine adjustment that brings the centroid of an aperture-limited image to the normal foveal axis.
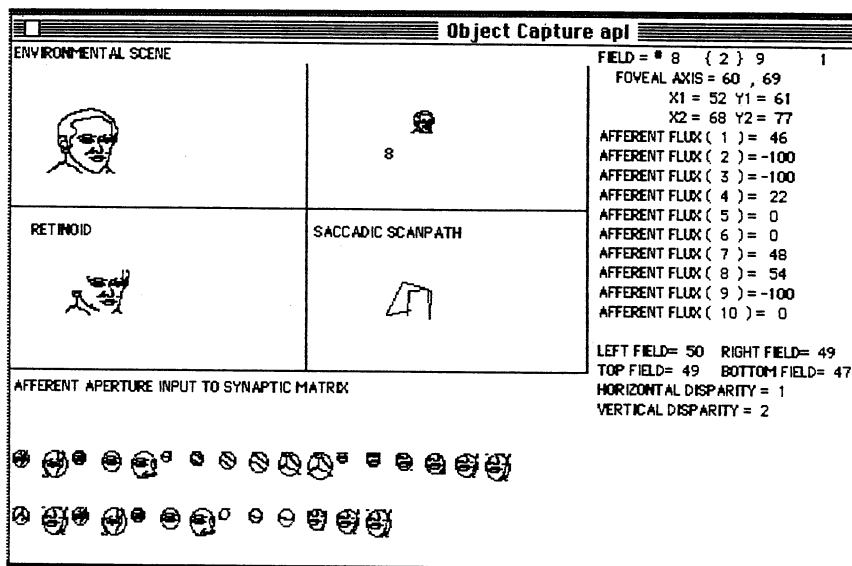
```
                                            Object Capture apl
ENVIRONMENTAL SCENE                          FIELD = # 8   { 2 }  9        1
                                                FOVEAL AXIS = 60 , 69
                                                    X1 = 52 Y1 = 61
                                                    X2 = 68 Y2 = 77
                                        8    AFFERENT FLUX ( 1 ) =  46
                                             AFFERENT FLUX ( 2 ) = -100
                                             AFFERENT FLUX ( 3 ) = -100
                                             AFFERENT FLUX ( 4 ) =  22
                                             AFFERENT FLUX ( 5 ) =   0
RETINOID              SACCADIC SCANPATH       AFFERENT FLUX ( 6 ) =   0
                                             AFFERENT FLUX ( 7 ) =  48
                                             AFFERENT FLUX ( 8 ) =  54
                                             AFFERENT FLUX ( 9 ) = -100
                                             AFFERENT FLUX ( 10 ) =  0

                                             LEFT FIELD= 50   RIGHT FIELD= 49
                                             TOP FIELD= 49   BOTTOM FIELD= 47
AFFERENT APERTURE INPUT TO SYNAPTIC MATRIX   HORIZONTAL DISPARITY = 1
                                             VERTICAL DISPARITY = 2
```

Figure 7.8
CRT display of object-parsing simulation. The stimulus, shown in the upper-left panel, is a three-quarter frontal view of a human head. The bottom panel shows the parts of the head that were parsed after each of eight saccadic fixations: (1) left eye, (2) right eye and bridge of nose, (3) right jaw, neck, and shoulder, (4) mouth and lower nose, (5) mouth and lower nose, (6) left eye, (7) right eye and bridge of nose, (8) mouth and lower nose. Left-middle panel shows the excitation pattern on the scene assembly retinoid. Notice the partial reconstruction of the head and face with the parsed features in proper spatial relation.

Remembering the Location of Objects

There are at least two ways in which the location of objects viewed in an extended environmental scene might be represented in memory for later recall. One is to reduce the size of the scene assembled in the retinoid system. Then it can be projected to the synaptic matrix as a complete spatial layout to be learned and evoked in the imaging matrix when the information is needed. The other is to register the 3-D coordinates of each object in the scene and learn them together with their associated objects. The first procedure would require an appropriate dimensional reduction of the retinoid output by operation of the size transformer (see figure 7.1) so that the entire scene can be captured (within the afferent aperture) by the synaptic distribution on a single filter cell. Although this approach might serve to represent the relative location of objects in 2-D space, location in
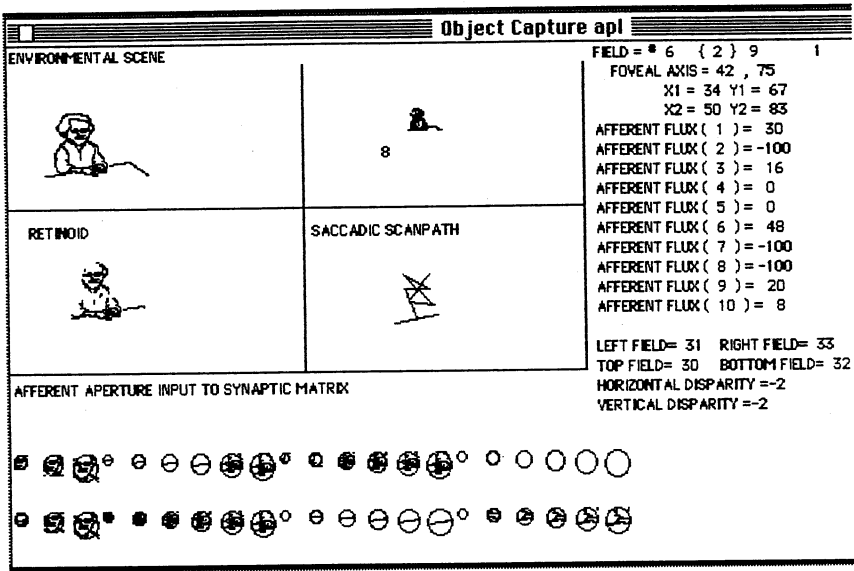
Figure 7.9
CRT display of object-parsing simulation. The stimulus, shown in the upper-left panel, is a person seated at a table. The bottom panel shows the parts of the scene that were parsed after each of eight saccadic fixations: (1) full face, (2) hands, (3) hands, (4) nothing (no centroid adjustment within tolerance), (5) full face, (6) hands, (7) edge of table, (8) right forearm. Left-middle panel shows the excitation pattern on the scene assembly retinoid. Notice the partial reconstruction with parsed features in proper spatial relation.

depth would not be conserved. This follows from the fact that excitation from the 3-D retinoid is a 2-D projection to the mosaic cells of the synaptic matrix. In the second approach, during recall, the imaginal representation of each object or its class cell token would be separately evoked in the imaging matrix, projected to the 3-D retinoid, and then translated in retinoid space to its original autaptic cell coordinates. In this case, because of the connectivity for size constancy in the 3-D retinoid, if the pattern of an object (rather than its token) is projected to the retinoid, its representation would diminish in size in accordance with its distance on the Z-axis

In this routine, it is assumed that focal attention is required to learn the location of an object. Thus, there is an excursion of the self-locus to the retinoid location of each object of interest before it is translated to the normal foveal axis. The model assumes that each autaptic cell in the 3-D retinoid sends out three collateral axons providing selective input to three groups of neurons that sense the planar location of

that autaptic cell on the X, Y, and Z axes respectively (if its level of activity is sufficiently high). Activity of autaptic cells at each of the successive target positions of the self-locus is increased by the added input from the self-locus (exogenous stimulation from a retinal pattern or endogenous stimulation from an object token, plus excitation from self-locus discharge). The augmented output from each of the targeted autaptic cells is sufficient to exceed the threshold of a discrete neuron in each of the three groups of labeled lines that sense the location of neuronal activity on the X, Y, and Z axes, respectively. These triplets of labeled lines constitute neuronal indexes to the approximate spatial location of each object or token that is attended to in the scene. Figure 7.10 shows how the position of the self-locus stimulates labeled lines for sensing the X, Y, and Z coordinates that define the retinoid locations of two separate objects. The cell triplets (C_X, C_Y, C_Z) then provide input to a synaptic matrix where object tokens are associated with their appropriate spatial coordinates (figure 7.11).
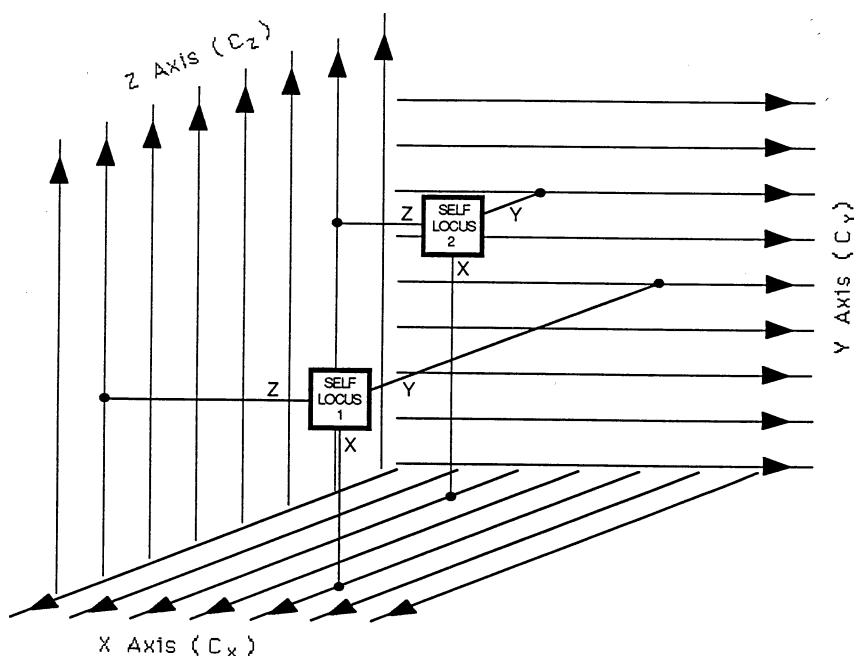
Figure 7.10
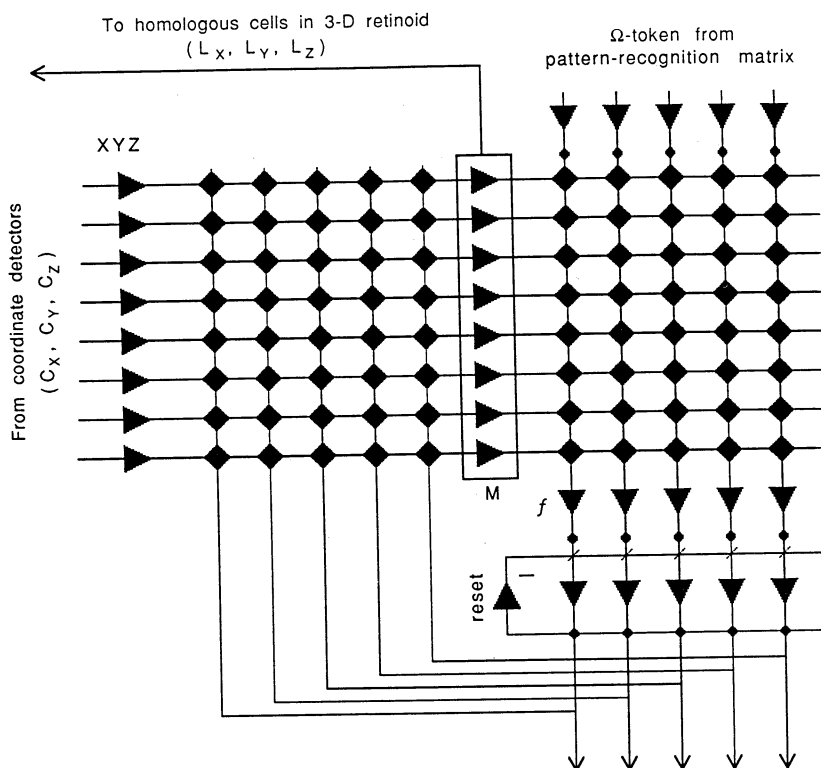Neuronal mechanism for sensing the location of self-locus excursions in retinoid space.

Figure 7.11
Synaptic matrix for associating object tokens with their spatial coordinates in retinoid space.

After the association between a token and its retinoid location has been learned, any evocation of that token alone will induce the discharge of three locator cells (L_X, L_Y, L_Z) via feedback to the 3-D retinoid from their homologous mosaic cells in the imaging matrix (figures 7.11 and 7.12). The function of the locator cells is to activate selected autaptic units at those coordinates within the 3-D retinoid that correspond to the location of each object of interest when the original scene was learned. When positional recall of this kind occurs, the firing of each object token is accompanied by a "brightening" (autaptic cell discharge) at its original location in the 3-D retinoid. This happens automatically because the output from any triplet of locator cells will summate fully and exceed threshold only in the autaptic cell that is at the retinoid coordinate defined and selected by the particular triplet L_X, L_Y, L_Z (figure 7.12).
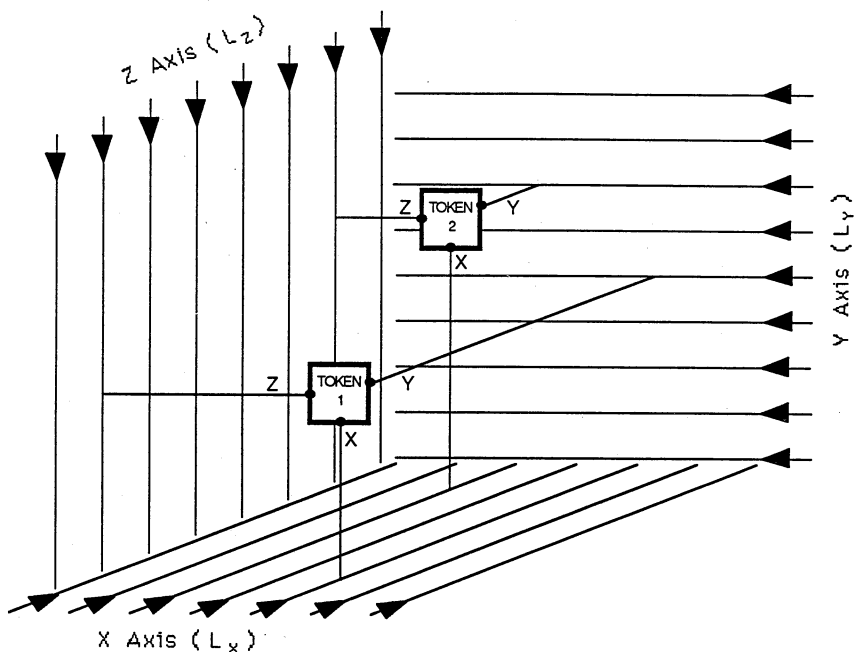
Figure 7.12
Neuronal mechanism for signaling the original location of an object in retinoid space.

A key assumption of this model—that the activity of retinoid cells at the egocentric coordinates of objects of interest in a visual scene will be selectively augmented— is supported by the experimental findings of Andersen, Essick, and Siegel (1985). These investigators report that single neurons in the inferior parietal lobule of behaving monkeys show enhanced spike activity in response to visual stimuli of standard intensity as a systematic function of the angle of gaze. Thus, there is an eye position–dependent tuning of cortical cells for locations in head-centered coordinate space.

Representing Spatial Relationships among Objects

How can the spatial relationships that exist among multiple objects in a scene be analyzed and discretely represented in the brain? How, for example, can we determine the relative distance between object A and object B? Is A above or below B or to the right or left of it? Is A nearer than B or farther away than B to an arbitrary point of reference in space? Quick solutions to such problems are essential for

effective action in response to immediate contingencies in the real world. And beyond immediate response, the simplest plans for traveling in an environment that presents multiple constraints on mobility require knowledge of such spatial relationships.

Determining the Position of an Object Relative to Another
Whereas a parallel excitation pattern over the 3-D retinoid can give a holistic neuronal-analog depiction of the spatial layout in a visual environment, a description of the particular pairwise relationships among objects in a scene requires a sequential analysis of the layout. In analyses of this kind, the neuronal triplet C_X, C_Y, C_Z that defines the spatial coordinates of one object must be held in working memory until it can be compared with the triplet that gives the location of the second object. This is accomplished by three separate pairs of autaptic cell buffers that are homologously ordered for C_X, C_Y, C_Z respectively. The corresponding autaptic cells within each pair of buffers have cross-inhibitory coupling. These buffer pairs serve as comparators that signal the relative position among objects on each axis in 3-D space (figure 7.13).

In the example given at the top of figure 7.13, both objects, as represented in their respective buffers, are in the left hemifield of the X-axis. The two autaptic representations are then translated to the right, each inducing a string of excited cells, until the translation of one of them is blocked by cross-inhibition from a leading active cell in the matched buffer. The leading representation, which cannot be blocked by cross-inhibition, continues its extension to the right until it fires the terminal cell in its buffer. When a terminal autaptic cell in either of the two buffers is activated, it discharges a terminal-sensing cell (T_1 or T_2). At this point, the activity of the translation driver D is squelched by an inhibitory input from the T cell that has fired. In the case shown, T_1 fires, indicating that object 1 is to the right of (leading) object 2 (taking object 2 as the reference) and object 2 is to the left of (lagging) object 1 (taking object 1 as the reference). There are similar buffer pairs for the Y and Z axes, respectively, which operate in the same way. In this fashion, the relative positions of each pair of objects that are the targets of focal attention are individually coded by the output of the T cells for each of the three spatial axes (left-right, above-below, nearer-farther).

Determining the Distance between Objects
The buffer comparators that analyze the relative position among objects also generate information about the distance between objects. In figure 7.13, after translation, object 2 occupied just one more au-
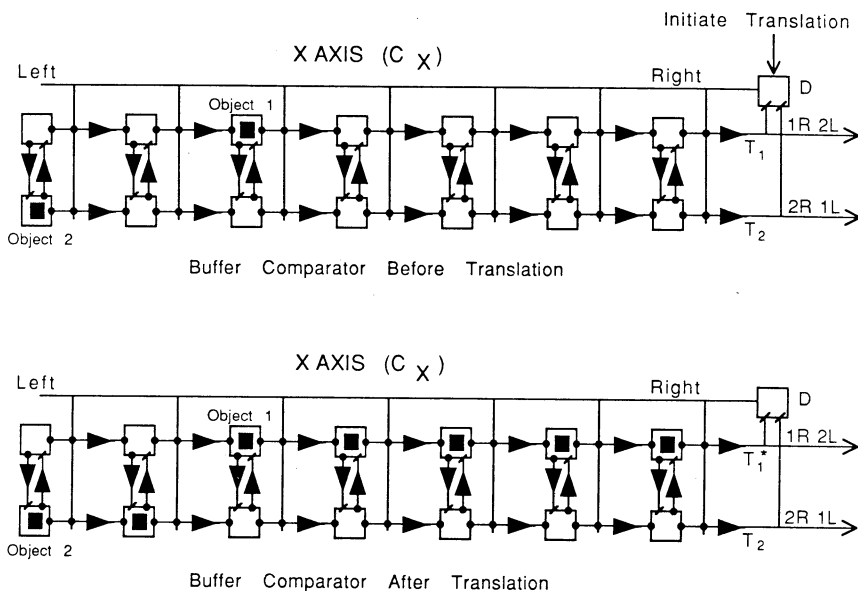
Figure 7.13
A buffer comparator for sensing and signaling the relative position among objects on the X-axis of retinoid space.

taptic cell than before translation. This follows from the fact that in the retinoid analog of the perceived scene, object 2 was separated from object 1 by a single autaptic cell. Had there been a greater distance on the X-axis between the objects, the number of autaptic units activated by object 2 after translation would have been proportionally greater. If object 2 were to the right of object 1, the state of the buffers would be exactly reversed. Thus, the number of active autaptic cells in the buffer that holds the spatial coordinate of the lagging object is an index to the distance between objects. The same principle operates in the buffer comparators for the Y- and Z-axes as well.

One way to capture this information on a labeled line is to sum the autaptic cell activity in the lagging buffer as input to a single "distance" cell for each axis. Figure 7.14 illustrates how this might be done. Output from the T cells is inhibitory, and, for each axis, the particular T cell that fires is just that terminal-sensing cell activated by the buffer holding the coordinate of the leading object. Activity in each of the buffer pairs is summed independently in two interneurons. The output from these, in turn, is summed in a single output cell. However, inhibitory collateral axons from T_1 and T_2 ensure that
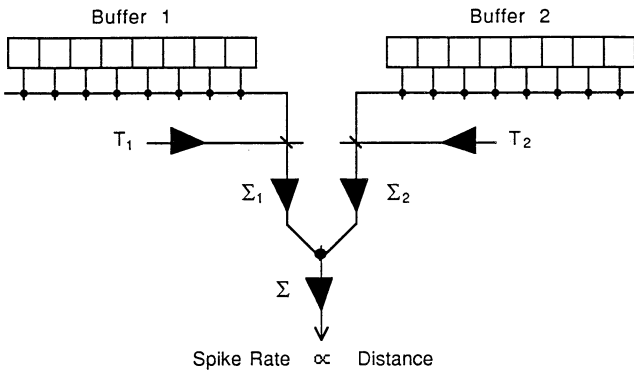
Figure 7.14
Neuronal mechanism for sensing and signaling the relative distance between two objects.

whichever of these is active will block input from the leading buffer, thus allowing only input from the lagging buffer (the desired source) to modulate the frequency of firing of the labeled line that signals the distance between objects. In this scheme, if a triplet is composed of the activity of just one autaptic cell in each of the lagging buffers for the X-, Y-, and Z-axes, the two objects are in contact (or at least so close as to be indistinguishable from the relationship of direct contact). In addition to signaling direct contact between objects, this kind of neuronal information can be used to guide the heuristic movement of the self-locus for tracing and registering the contours of objects or aspects of the environment. The ability to trace and learn significant environmental contours such as barriers and pathways permits the brain to store simple but useful schematic representations of a complex sensory world.

Characterizing the Movement of an Object
We have seen how the relative spatial relationships among different objects in a perceived environment can be analyzed and represented by neuronal mechanisms. The spatial coordinates of a single object at different times can be analyzed and represented in a similar fashion. Instead of comparing a pair of objects, two successive locations of a given object can be captured and compared. In this way, its movement and the direction of movement can be represented.

One might question whether there would be any advantage to having a separate mechanism of this kind when we know that there are motion detectors at lower levels of the visual system (Barlow,

Hill, and Levick 1964). It should be recalled, however, that in the mechanisms under discussion, focal attention (excursion of the self-locus) is required to register the spatial coordinates of objects of interest. This suggests that one advantage of these mechanisms over lower-level motion detectors that are indifferent to stimulus content is that goal-related computations of movement can be selectively performed in an environment that might have a considerable amount of incidental motion. In this account, a person could be well aware of a variety of objects moving about in the immediate environment yet be able to isolate and determine in particular the course of those self-selected objects among them. A recent investigation of motion detection cells in monkey cortex presents some evidence to support such a mechanism. Saito and associates (1986), recording from single cells in the superior temporal sulcus, discovered, near cells that were selective for the direction of image motion but not image content, a cluster of directionally selective cells that responded only to movements of real objects (face, hand). When dot patterns, large spots, or other objects (that presumably did not engage focal attention) were moved in the preferred direction of these cells, they evoked no significant response.

Lexical Assignment
Outputs from the mechanisms described are neuronal tokens of spatial relationships between different objects in the environment or between successive positions of single objects. It is assumed that they provide input to a lexical assignment matrix and are mapped to appropriate words in a language system in the same way that tokens of specific objects are mapped to words. Just as the stimulus provided by a kind of small, furry animal can induce the noun *cat*, so can the stimulus provided by a kind of spatial relationship between objects induce the appropriate adverb, adjective, or preposition (*above* or *below*). Moreover, by the same kind of neuronal architecture for back-mapping that evokes images of specific objects when their lexical tokens are fired, the discharge of a lexical token of a spatial relationship can evoke an analog representation of the relationship that is named.

Cognitive Maps

Knowledge of a particular environment entails more than having an internal representation of the objects in it, their locations, or their individual properties and affordances. It also includes a representation of potential routes—unimpeded pathways that allow personal

access to objects and locations of interest. Information of this kind can be learned in the course of experience and stored as a list of propositions about a particular locale, or it can be learned and stored as a schematic analog representation—a cognitive map of a significant place in the world, to be recalled in a single frame and utilized for coping within the environment that is represented. The retinoid system and the analytic mechanisms provide sufficient neuronal means for constructing such cognitive maps. The ability to project heuristic excursions of the self-locus along the routes represented in a recalled map enables one to choose among alternative plans for moving from an initial starting point to successive target regions (or objects) in the associated environment.

In previous discussions of the retinoid system, the normal foveal axis was taken as the reference for direction of translation. This established an egocentric frame for representing the environment. Cognitive maps, however, require an object-centered frame (Marr 1982, Pinker 1984) or geo-centered frame in order to provide useful information for maneuvering in the environment. This can be accomplished by translating a feature (a significant part) of an object or aspect of the environment to a point on the normal foveal axis in retinoid space, fixing its token at this location, and then using this coordinate as the reference origin for controlling subsequent spatial translations. For example, the entrance to a room could be the fixed origin in retinoid space and would be taken as a local region centered on the normal foveal axis at the frontal plane of the room. Front, back, left, and right would be directions in relation to the entrance door. Looking from the doorway at a desk in the front-right corner would entail a direct right excursion of the self-locus. Looking at the same desk from the wall opposite the door would entail an excursion of the self-locus from a back-of-the-room position in retinoid space to a front-left location. Thus, the coordinates of object (room) features are stabilized, and the self-locus is translated with reference to these fixed coordinates.

Figure 7.15 shows the CRT display of a 2-D simulation that illustrates how heuristic excursions of the self-locus during a walk in an unfamiliar building can create an autaptic cell map of the corridors by which particular offices can be reached. The number of autaptic cells that are activated by alternative routes of the self-locus is proportional to the relative distance that must be traveled in the course of each route. Choosing the route with the least total activation ensures that one has selected the shortest route.

In this example, a person (S) is assumed to have exited from a subway station and is facing a large building with two entrances,

File Edit Search Run Windows

```
                                    Cog Map
NUMBER OF TARGETS = ? 6                              START = 306  - 83
                                                    GOAL 1  =  266  - 69
                                                    GOAL 2  =  246  - 63
                                                    GOAL 3  =  245  - 33
                                                    GOAL 4  =  268  - 34
                                                    GOAL 5  =  304  - 79
A        START              C        START


   3                           3
   2                           2
B        START         D        START


   ROUTE 1 =  213                       UP =  25      DX 6  = -2
   ROUTE 2 =  128                       LEFT =  41    DY 6  = -2
                                        DOWN =  22
                                        RIGHT =  40
```
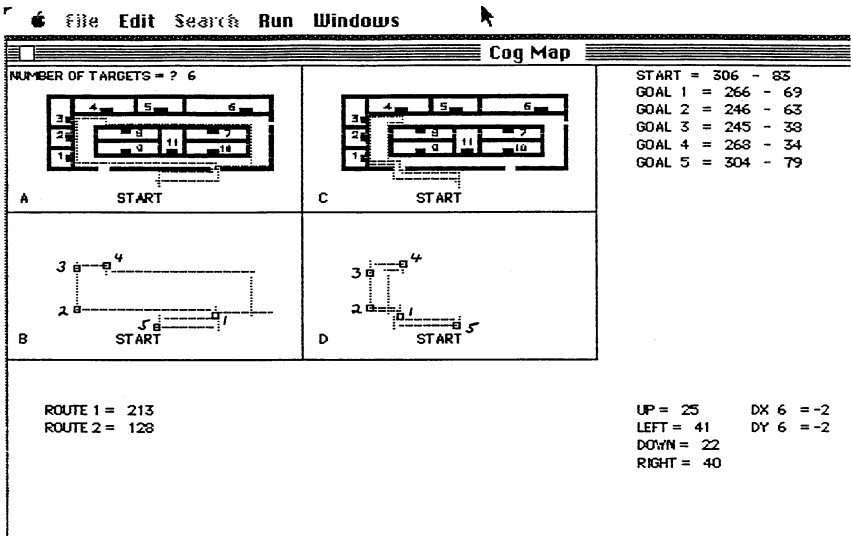
Figure 7.15

CRT display of cognitive map simulation. Panels A and C show the floor plan of an office building. Dotted lines in A and C represent two different routes taken by an individual (S) who transacts business in offices identified as 1, 3, and 4 on two separate occasions. Dotted lines in panels B and D represent the self-locus excursions made in object-centered retinoid space while S walks the routes shown in A and C. Small squares in B and D represent goal regions. Numbers beside squares represent the sequence in which the goals are set. The autaptic cell excitation patterns induced by excursions of the self-locus as shown in B and D are each learned holistically as two possible routes to be taken when transacting business in this building (cognitive maps). The total number of cells traversed and activated by the self-locus over each route is independently summed (bottom panel) and represents the distance traveled in completing each of the routes. This sum is physiologically represented in the relative discharge frequency of a cell that integrates total output from the self-locus retinoid and serves to label each route by its overall distance.

one to the left and the other to the right. The two entrances are approximately equidistant from the subway exit. There are three offices (rooms 1, 3, and 4) in which S must transact business and then return to the subway station for the trip home. S is assumed to be able to recognize and remember whatever is in his visual field (walls, doorways, signs, office numbers).

On the first visit (figure 7.15, panels A and B), the locations of the offices are unknown. S arbitrarily chooses the right-hand entrance because neither entrance is appreciably closer to the subway exit than the other. This entrance is the first goal to be reached and is fixed as the reference locus for the object BUILDING(n). Entering the building,

S is immediately confronted by a long corridor running left and right. The office directly across the entrance (room 10) is not one to be visited. At the far left end of the corridor, however, is room 1, one of the three in which business must be transacted. At this point, none of the other relevant offices is in sight, and room 1 becomes the second goal. Before S walks to room 1, however, the self-locus has already made an excursion to this target region at the end of the corridor in the left-hand part of retinoid space. Thus, the tokens of goal 1 (entrance) and goal 2 (room 1), together with the direction and distance of the self-locus excursion between the two goals, are roughly captured in the retinoid system by the spatial pattern of autaptic cell activity and the number of cells fired (traversed). In this process, each change in direction is determined by change in angle of gaze, and the distance of each leg of the excursion is determined by target location on the Z-axis.

When S reaches room 1, a glance to the right reveals the third goal, room 3, at the end of a corridor extending from the front to the back of the building with reference to the entrance (goal 1). A self-locus excursion is made to that homologous region of retinoid space (in the object-centered frame). Similarly, when S is at room 3, the last office that must be visited (room 4) can be seen. This is goal 4, and the self-locus is translated to the coordinates of this target. After all business has been transacted, S must find an exit from the building and return to the subway station (goal 5). S continues down the corridor where room 4 is located, turns right at the end of the corridor, turns right again, and finds the doorway through which he entered (the first goal). S leaves the building through the same doorway and goes to the subway station (goal 5). The route S traveled in this episode (route 1) is traced by the excitation pattern of the self-locus in its corresponding excursion within object-centered retinoid space (panel B). This pattern is projected to a synaptic matrix and stored as a cognitive map that can be imaged to serve as a possible guide on future visits to the same building.

A second visit to the same offices is shown in panels C and D of figure 7.15 (route 2). In this case, S arbitrarily chooses the left entrance to the building and goes to rooms 1, 3, and 4, as in the first visit. After leaving room 4, however, S turns right, retraces his path, exits through the same door by which he entered the building, and returns to the subway station. The number of autaptic cells traversed and discharged in routes 1 and 2, respectively, represents in a rough way the relative distance traveled for each route. Route 2 is clearly shorter (128 versus 213 autaptic units) and will be the cognitive map

followed when the same transactions must be undertaken in the future.

As S walks through the building for the first time, he may note, learn, and relate incidental aspects of the environment to the route traveled. Information of this kind can provide a basis for planning more efficient behavior in the same setting on another occasion. In the example, S made the decision to adopt a particular travel route after he actually followed the route and then compared the result with an earlier course of behavior. An alternative strategy would have S consult (image and analyze) the cognitive map created on the first visit to the building to determine the most efficient route for going to each office. Assuming that he noted and remembered the building entrance close to room 1, all of the information needed is given in the retinoid map of route 1. Before a second trip is undertaken, heuristic excursions of the self-locus (tracing the contours of the map) could reveal that route 2 is better and should be taken as the plan for the next visit.

During each visit, two distinct and parallel representations are constructed in separate registers of the retinoid system: the sequence of visual stimuli observed from S's successive viewpoints and integrated in the scene assembly retinoids and the autaptic cell trace of S's heuristic self-locus as it precedes S's actual motion from each successive goal to the next in an object-centered retinoid space. This trace composes a cognitive map of locomotion affordances among significant regions within the object BUILDING(n).

Given a sufficient level of arousal, the two kinds of retinoid representations can be learned in separate synaptic matrices: one receiving projections from the scene assembly register and the other receiving projections from the register that captures the cognitive map. Because the different object-centered coordinates of the map are temporally correlated with particular visual scenes during learning, class cell tokens of these coordinates in the BUILDING(n) map can be synaptically coupled to class cell tokens of successive scenes (in the same way that lexical tokens can be coupled to visual stimuli). Following the map in imagination (tracing its contour by excursion of the self-locus) will evoke a sequence of self-locus coordinate tokens, which, in turn, can evoke visual images of the sequence of scenes that were originally experienced in BUILDING(n). Alternatively, recalling the class cell token of a particular scene will evoke its coordinate location on the cognitive map of BUILDING(n).