REVIEW

Kuniyoshi Sakai,[1] Yuji Naya, and Yasushi Miyashita

Department of Physiology
School of Medicine
University of Tokyo
Bunkyo-ku, Tokyo 113, Japan

Neuronal Tuning and Associative Mechanisms in Form Representation

Abstract

We examine the hypothesis that the form representation in the anterior inferotemporal (AIT) cortex is acquired through learning. According to this hypothesis, perceptual aspects of the temporal association area are closely related to its visual representation, in that the response selectivity of AIT neurons can be influenced by visual experience. On the basis of the neurophysiological evidence, we summarize two neuronal mechanisms that subserve the acquisition of form selectivity in AIT neurons. The first mechanism is neuronal tuning to particular stimuli that were learned in a cognitive task. The second mechanism is association, with which relevant information can be retrieved from other stored memories. On the grounds that long-term memory of objects is acquired and organized by at least these two neuronal mechanisms in the temporal association area, we further present a model of the cognitive memory system that unifies perception and imagery.

Background

The inferotemporal (IT) cortex has been proposed to be the memory storehouse in object vision (Gross 1973; Mishkin 1982; Sakai and Miyashita 1993a). Along the visual pathway from the primary visual cortex (V1) to the anterior inferotemporal (AIT) cortex, both the receptive field size and the complexity of neuronal processing increase (Gross 1992; Van Essen et al. 1992). Consequently, IT neurons respond selectively to complex forms such as hands, faces, and computer-generated forms (Fig. 1). A recent neurophysiological study suggests that AIT neurons in adult monkeys are subject to tuning mechanisms for complex features of visual stimuli through lifelong sensory experience and learning (Sakai and Miyashita 1993b, 1994a).

FORM REPRESENTATION

Primate vision has a remarkable ability to recognize a variety of similar faces or objects. This ability suggests that form information is perceptually organized so that it enables fine discrimination of faces or objects. One recent psychological study demonstrates that even the brightness percept is influenced by the perceptual organization of high-level properties such as depth and form (Adelson 1993). In this review we use the term *form* for the geometry of an object's overall structure. We believe that form is an indispensable concept in understanding mechanisms of object recognition, because form directly represents an object's entity and enables its recognition.

There still remains a controversial question whether form perception
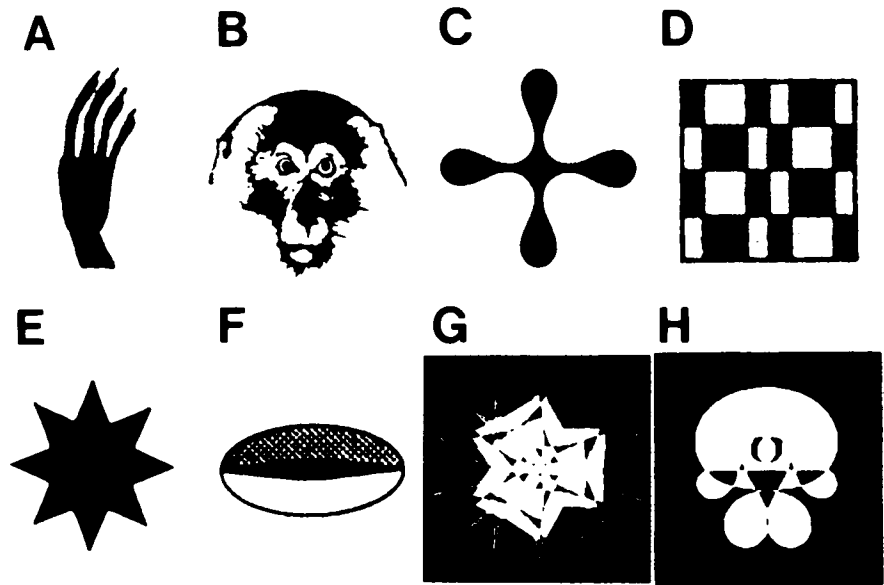
[1]Corresponding author.

Figure 1: Examples of the screening set used in testing the responses of IT neurons. One typical visual stimulus is shown for each screening set. (*A*) The shadow of a monkey hand (Gross et al. 1972); (*B*) a monkey face (Desimone et al. 1984); (*C*) a simple FD (Schwartz et al. 1983); (*D*) a two-dimensional Walsh pattern (Richmond et al. 1987); (*E*) a star form (Tanaka et al. 1991); (*F*) a pattern decomposed from the lips of a face (Fujita et al. 1992); (*G*) a fractal pattern (Miyashita and Chang 1988; (*H*) a complex FD (Sakai and Miyashita 1991).

occurs as overall identification or as a synthesis of structural components. Some theories of object recognition have proposed bottom-up determination of an object's components and subsequent matching of the arrangement of components with a memory representation (Marr 1982; Pentland 1986; Biederman 1987). The properties of face-selective cells found within the superior temporal sulcus (STS) may be in agreement with these theories (Perrett et al. 1987). However, there is psychological evidence that object features (not specifically related to parts) are matched directly with such overall features stored in long-term memory (Cave and Kosslyn 1993). Because this problem is inseparable from the nature of form representation itself, we will deepen this discussion in the section Is Form Representation 2D Viewer-centered or 3D Object-centered? (below) after presenting more relevant data.

FORM SELECTIVITY AND
ORIENTATION SELECTIVITY

Whereas the orientation selectivity of V1 and V2 neurons has been well characterized (Hubel and Wiesel 1959; Bonhoeffer and Grinvald 1991; Blasdel 1992), the form selectivity of IT cells has not been thoroughly studied, owing in part to the complexity of form features (Schwartz et al. 1983; Desimone et al. 1984; Tanaka et al. 1991). When recording a V1 cell, its orientation preference and selectivity can be systematically determined. Orientation is of one-dimensional parameter ranging from 0° to 180°. In contrast, the range of form variety is almost infinite. Moreover, no critical parameters are known for the specification of general forms.

If the form analysis in vision is hierarchical and bottom-up, then some aspects of the form selectivity of IT cells reflect the constraint in the

early visual processing. Besides, if this visual processing is based on computational logic, then there must be some principles that guide the generation of the form selectivity. In this article we propose the guiding principles of neuronal tuning and association in the AIT cortex. Our working hypothesis is that learning of a form or repeated exposures to a form produces neurons that selectively respond to that particular form. This idea bears some relation to prior conjectures for the experimentally controlled picture-selective responses (Miyashita 1988) and the modified responses of face-selective cells (Rolls et al. 1989).

FINE-FORM SELECTIVITY

The conventional method for determining the response selectivity of a single IT neuron utilizes a *screening set* of various object forms. Because experimenters are not able to test every possible form, they must prepare a convenient screening set with many visual stimuli. Figure 1 shows some examples of screening sets utilized so far. This kind of screening set can specify global form selectivity. If a recorded neuron responds to at least one of the forms in a screening set (responsive), but not equally to all of them (selective), the global form selectivity of this cell can be characterized further.

One may mistakenly conclude that the most effective form in a screening set is the optimum stimulus for a recorded cell. Suppose that we are now using the screening set of various hand forms. Even if a recorded cell responds best to a particular monkey's hand (shown in Fig. 1A) among this set, there is no guarantee that this hand is the optimum stimulus, because of the possibility that other untested hand forms may elicit much stronger responses. Moreover, an untested form with a global feature similar to the hand in Figure 1A may be the optimum stimulus. One way to overcome the limitation of using a screening set is the analysis of *fine-form selectivity*, which provides better resolution to discriminate among effective stimuli with similar features. In other words, the test of a cell's form selectivity should be performed in two steps. First, test the cell's responses with a broad screening set. Second, test them more finely by preparing forms similar to the effective stimuli selected in the first step. The analysis of fine-form selectivity becomes particularly important in searching for memory traces of specific forms, which could be acquired through learning experience.

PAIR–ASSOCIATION TASK

Figure 2 shows a screening set that we have used. These pictures were sorted randomly into pairs, numbered from 1 and 1' to 12 and 12'. The pair combinations were fixed throughout the experiments. This screening set is not only utilized for testing a cell's global form selectivity as noted above but also for testing how these stimuli are represented by AIT neurons after they are memorized. To impose the acquisition of the screening set as long-term memory, we trained monkeys (*Macaca fuscata*) in the pair–association (PA) task (Sakai and Miyashita 1991; Murray et al. 1993). In each trial of the PA task, a cue stimulus is presented at the center of the video monitor for 0.5 sec. After a delay period of 5 sec, two stimuli, the paired associate of the cue (correct choice) and a distractor (incorrect choice), are shown. The monkey obtains a fruit juice reward for touching the correct one within 1.2 sec. This task paradigm can reliably demand the learning of visual stimuli, because monkeys cannot select a paired associate correctly without memorizing and recalling pair combinations.
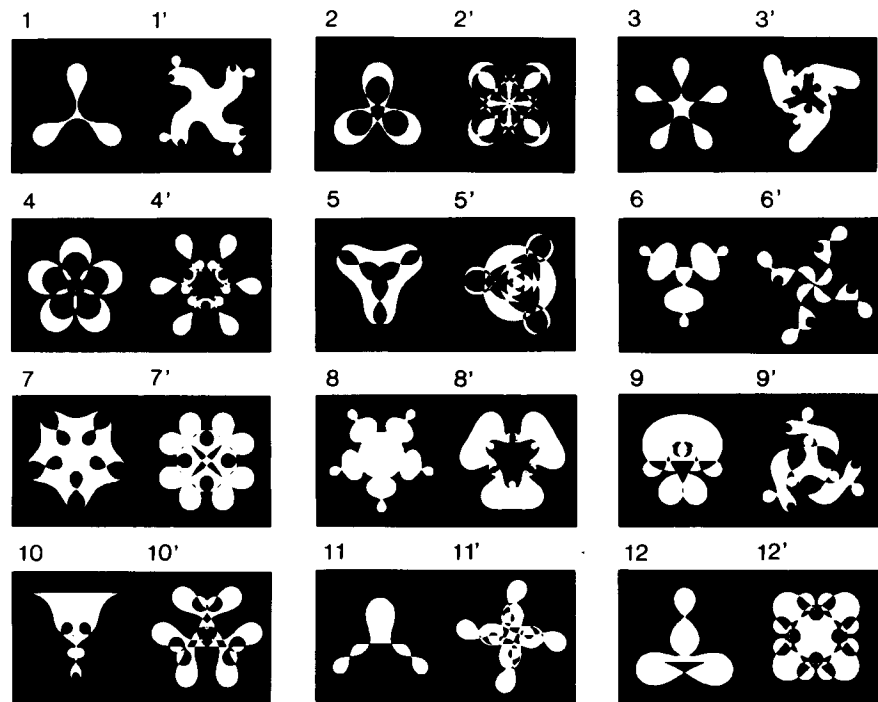
Figure 2: Twelve pairs of FDs (1 and 1' to 12 and 12') for stimuli in the PA task. When one member of each pair is shown, trained monkeys can retrieve and select the other member of the paired associates.

It should be noted that the PA task cannot be solved by employing short-term or working memory within a single trial. Instead, the PA task is essentially the memory *recall* task, which explicitly demands the memory retrieval and thus generation of images from the long-term memory. Therefore, memory components of the PA task present a sharp contrast to those of the delayed matching-to-sample (DMS) task. In the DMS task, the subject indicates whether a test stimulus matches a sample stimulus shown previously (Gaffan 1985). When the stimulus set size is small and thus the same trial is repeated, working memory is mainly involved in the DMS task. When the stimulus set size is large and thus each trial is unique, recognition memory is also involved, because it is possible to indicate whether a test stimulus has appeared or not without employing working memory. We have to take these points into consideration in interpreting the results of behavioral and physiological studies.

The Neuronal Tuning Mechanism in the AIT Cortex

PARAMETRIC TRANSFORMATION AND GEOMETRIC TRANSFORMATION

We have adopted computer-generated Fourier descriptors (FDs) as visual stimuli (Sakai and Miyashita 1991). FDs are appropriate in testing fine-form selectivity because they are specified by a set of parameters: harmonic amplitude (A_k) and phase angle (α_k), where $k = 1, 2, \ldots$, is each term's ordinal number in a Fourier series (Zahn and Roskies 1972). The number of free parameters for each pattern shown in Figure 2 is only

one or three, which corresponds to one or two sets of $[A_k, \alpha_k]$, respectively. (Note that one of the phase angles is not independent and affects only the overall rotation.) A slight alteration of one FD parameter from its original value produces a very similar form. This manipulation of an original pattern is called *parametric transformation* (Fig. 3A). The similarity of forms with slightly altered FD parameters is validated by the fact that the position of each point in the drawing plane is a continuous function of FD parameters (some examples are shown in Fig. 4, below). The parametric transformation should be regarded as a method for changing forms systematically in testing the fine-form selectivity. One previous study of IT cells reported that the neuronal responses to forms produced by combining two FD terms cannot be predicted from the responses to each component FD (Albright et al. 1985). However, this procedure leads to *global* alteration of forms, which is independent of the *local* parametric transformation.

There is another type of manipulation, *geometric transformation*, in which an original pattern was rotated about its center with various rotation angles or was reflected as a mirror image (Fig. 3A). The geometric transformations show different views of the same form, whereas the parametric transformations show different forms with similar global features. It has been reported that IT cells and face-selective STS cells show invariant responses over changes in stimulus size, contrast, orientation, and retinal location, thereby activating a different set of V1 neurons (Schwartz et al. 1983; Rolls and Baylis 1986; Miyashita and Chang 1988). Furthermore, the form selectivity of IT cells did not vary
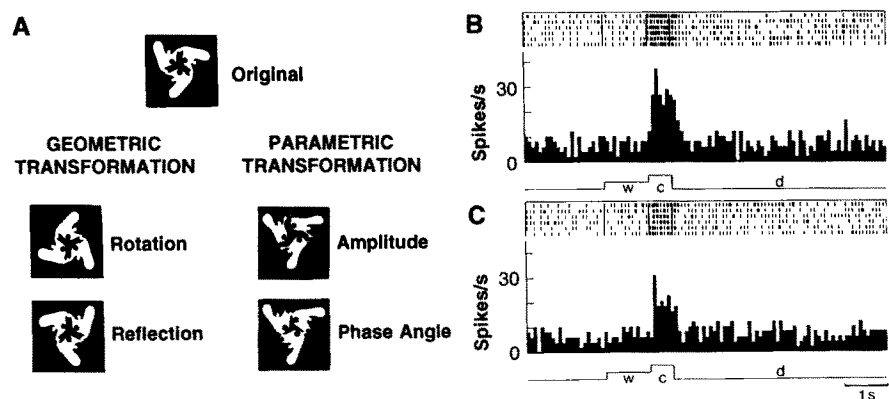

Figure 3: (A) Two types of transformations of FDs. Geometric transformations include rotation and reflection of an original pattern. Examples of rotation and reflection are shown as *upper left* and *lower left* patterns, respectively. Parametric transformation is a slight alteration of one FD parameter. Examples of harmonic amplitude and phase angle alterations are shown as *upper right* and *lower right* patterns, respectively. The actual images on a video monitor were yellow monochrome against a black field. (B) Neuronal response to the cue stimulus of the original pattern shown at the *top* in A. Rastergrams of single-cell discharges in each trial (*top*) and spike-density histograms (*bottom*) are shown. Bin width: 80 msec. (C) Neuronal response to the cue stimulus of the pattern shown at *lower right* in A. The neuron illustrated in B showed weaker responses to this parametrically transformed pattern than to the original pattern during the cue period in each trial. These collected trials were originally separated by intervening trials of other cue stimuli and were sorted by off-line computation. (w) Warning period; (c) cue period; (d) delay period.

with visual cues used to define the form, such as luminance, texture, and relative motion (Sáry et al. 1993). Consistent with these findings, the responses of AIT neurons to geometrically transformed patterns were, on average, equal to the original stimuli (Sakai and Miyashita 1994a).

FORM DISCRIMINATION METHOD

The original 24 patterns in the screening set (Fig. 2) correspond to learned forms. Patterns with parametric transformation or geometric transformation are unlearned transforms, because they were not shown to the monkey before recording sessions. Both types of transforms were treated in the same way in the task. The geometric transformation can serve as a *sham manipulation* to incorporate transformed patterns. The PA task forces the monkey to discriminate between the learned forms and unlearned transforms by imposing a new association. For each recorded neuron, we select an original pattern (e.g., 3′ shown in Fig. 2 and Fig. 3A, top) that elicits the strongest or the second-strongest response among the 24 patterns during the cue period. Pattern $\hat{3}'$, transformed from 3′, is used to temporarily replace one original pattern (e.g., 9′) that is randomly selected for each recording session. This procedure has two advantages. First, a new association between two patterns ($\hat{3}'$ and 9 in this case) is imposed along with the original pair combination (3′ and 3) to discriminate $\hat{3}'$ from 3′. Second, a neuronal response to 3′ can be recorded and compared with a response to $\hat{3}'$ in the same block of trials.

EVIDENCE OF NEURONAL TUNING

We recorded extracellular discharges of each single neuron in the AIT cortex, as shown in a previous report (Sakai and Miyashita 1991). Of 474 isolated neurons in three hemispheres, 89 cells reproducibly showed form-selective responses during the cue period. The most effective stimuli for the 89 cells covered all patterns in the screening set. Of the 89 cells, 40 were tested for transformed patterns in stable recording conditions. The intensity of the strongest neuronal responses to the screening set ranged from 25.0 to 77.5 Hz (44.9 ± 14.2 Hz, mean $\pm$ s.d.; $n = 40$). Figure 3, B and C, shows data from one neuron with form-selective responses. The critical comparison is the neuronal response to the learned forms versus unlearned transforms during the cue period. Although both the learned original pattern (Fig. 3B) and the unlearned pattern with a slightly altered phase angle (Fig. 3C) elicited the cue response, the former had a significantly stronger effect than the latter.

Because two or more transformed patterns were tested in each of 10 cells, we analyzed 54 cases of transformations. The original FD pattern shown in Figure 4 has two amplitude parameters (A_4 and A_8) that can take continuous values. We prepared four patterns with a slightly altered value of A_4 or A_8. It can be readily noted that the global features of these altered patterns are strikingly similar (Fig. 4). All of the transformed patterns elicited significantly weaker responses than the original pattern. This observation was confirmed further by responses of other neurons. The parametric transformation resulted in a significantly weaker neuronal response for most cases. Moreover, we did not observe any case in which responses to transformed patterns were stronger than those to the original patterns. The difference between the effects of parametric and geometric transformations on neuronal responses was statistically
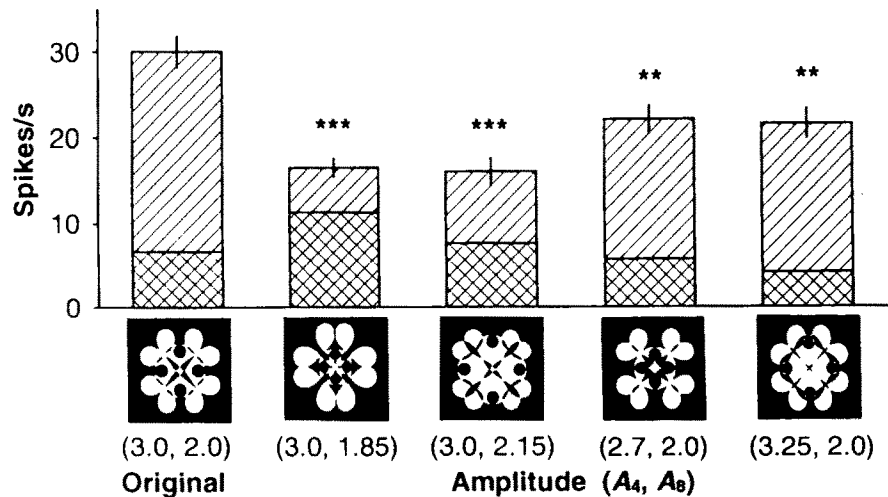
Figure 4: Responses of an exemplary neuron for the parametric transformation. The original pattern shown is specified by two harmonic amplitudes and two phase angles ($A_4 = 3.0$, $\alpha_4 = \pi/2$, $A_8 = 2.0$, $\alpha_8 = \pi/2$). The set of two numbers attached to each pattern denote its amplitude parameters (A_4, A_8). In each transformed pattern, one of these parameters is slightly altered from the original value. Note the close similarity between these four transformed patterns and the original. Histograms show mean discharge rates (mean ± s.e.m.) for each cue presentation of the original or transformed patterns. The crosshatched area in each bar corresponds to the spontaneous discharge rate. This neuron exhibited optimal responses to the original pattern rather than to the transformed patterns. These data were taken from the first five trials in which the original or transformed pattern was used as a cue stimulus. (**) $P<0.01$; (***) $P<0.001$: according to modified t-test (Snedecor and Cochran 1989).

significant ($P<0.01$, $n = 54$, Mann-Whitney U test) (Sakai and Miyashita 1994a).

Because of the following reasons, we can exclude the possibility that the weaker response to unlearned transforms is attributable to purely sensory reasons. First, every tested transform was derived from one of the original patterns, which elicited the strongest or the second-strongest response. Therefore, these transforms are likely to produce equally strong responses as the original pattern, obviously stronger than other ineffective original patterns. Furthermore, it is impossible to predict beforehand the relative effectiveness between the original pattern and transforms that derived from it, because original patterns have no special form attributes and their FD parameters were selected randomly. Second, a cell's preference among the screening set should not be confused with fine-form selectivity among transforms, because the range of form variety in the screening set is incomparably wider than that in transforms. Therefore, the learning-dependent *fine*-form selectivity cannot be explained by the *global* form selectivity for the screening set.

Furthermore, the weaker neuronal response caused by the parametric transformation is not the result of extraretinal factors such as attention. The geometric transformation did not produce a comparable reduction in neuronal responses even with the use of the same procedure to impose a new association with a transformed pattern (Sakai and Miyashita 1994a). This significant difference should be ascribed to the parametric manipulation of forms. These results suggest that AIT neurons are subject to tuning mechanisms for particular forms in a learning process.

TUNING AS LOCAL MAXIMA

Suppose the parameter space of FDs, in which each axis corresponds to one FD parameter (x_i) (e.g., amplitude A_4 in Fig. 4). Because every FD pattern is specified by a set of parameters $(x_1, x_2,\ldots, x_n)$, each pattern corresponds to a single point in this parameter space. Note that the 24 points of original patterns in the screening set are chosen randomly and sparsely. The parametric transformation corresponds to a local perturbation $(x_1^\circ,\ldots, x_{k-1}^\circ, x_k^\circ + \Delta x_k, x_{k+1}^\circ,\ldots, x_n^\circ)$ around the point of an original pattern $(x_1^\circ, x_2^\circ,\ldots, x_n^\circ)$, where $|\Delta x_k|$ is less than half the minimum separation of the points of original patterns (Fig. 5). We assume a continuous mapping from each point in the parameter space to a single-cell discharge rate $f(x_1, x_2,\ldots, x_n)$, that is, there are no break points in the response curve. As emphasized in the section Fine-form Selectivity (above), there is no guarantee that the point of an original pattern $(x_1^\circ, x_2^\circ,\ldots, x_n^\circ)$, which elicited the strongest response among the screening set, gives a local maximum of $f(x_1^\circ,\ldots, x_{k-1}^\circ, x_k, x_{k+1}^\circ,\ldots, x_n^\circ)$ (Fig. 5). Nevertheless, the parametric transformation always caused a

Neuronal responses
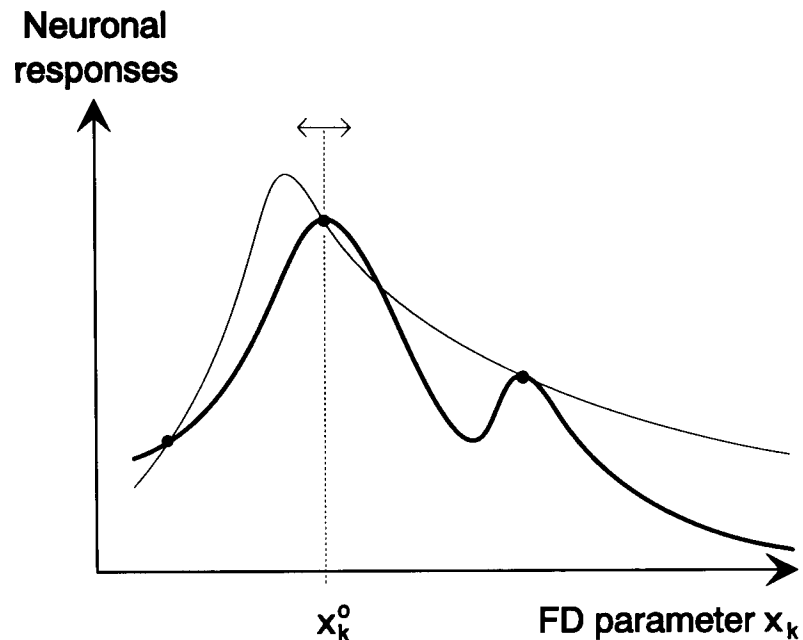

Figure 5: Neuronal tuning as a local maximum of responses. Schematic single-cell response curves $f(x_1^\circ,\ldots, x_{k-1}^\circ, x_k, x_{k+1}^\circ,\ldots, x_n^\circ)$ are shown as the function of one FD parameter x_k, where $(x_1^\circ,\ldots x_k^\circ,\ldots, x_n^\circ)$ is the point of an original pattern that elicited the strongest response among the 24 original FDs in the screening set. These 24 patterns provide only discrete points (three of them are shown) from which the response curve must be estimated. Two examples of interpolated graphs are shown as thin and thick curves. Here, we assume that the *middle* and the *right-most* points give strong responses, whereas the *left-most* point meets the spontaneous discharge level. The parametric transformation corresponds to a local perturbation around x_k°, as denoted by the double-headed arrow. The local change of the response curve in the double-headed arrow region may be a monotonic decrease as shown by the thin curved line, where half of the time the parametric transformation would cause an increase in neuronal responses. The experimental results indicate that the parametric transformation causes a decrease in neuronal responses for any tested points, as shown by the thick curved line where x_k° gives a local maximum. Accordingly, neuronal responses should be optimally tuned to the original patterns.

decrease in neuronal responses, which indicates that any tested points of original patterns give local maxima in responsiveness. Accordingly, neuronal responses should be optimally tuned to the original patterns. This neuronal tuning cannot be explained by learning-independent innate selectivity, because the original patterns in the screening set were selected randomly in the form space.

Here, we emphasize the importance of the analysis of fine-form selectivity. If the range of the parametric transformation were too large, it would be impossible to study neuronal tuning when there are several local maxima in the form space. Therefore, it was crucial to utilize the *local* parametric transformation to examine the fine-form selectivity.

TUNING VS. REPETITION—SUPPRESSION

In the DMS and serial recognition tasks for monkeys, there are several reports that IT cells show significantly weaker responses to "familiar" visual stimuli than to "novel" stimuli (Baylis and Rolls 1987; Miller et al. 1991b; Riches et al. 1991; Fahy et al. 1993; Li et al. 1993). The distinction between novel and familiar is such that if a novel stimulus is presented repeatedly and is held in memory, it becomes familiar. Although these IT cells may function as adaptive filtering that preferentially passes information about new stimuli (Miller et al. 1991b; Desimone 1992; Li et al. 1993), this repetition–suppression effect is closely coupled with at least three factors. The first one is habituation that causes a decrease of neuronal responses (Miller et al. 1991a). The second one is attentional enhancement to novel stimuli, which leads to stronger responses. The third one is priming that improves identification of perceptual objects (Tulving and Schacter 1990). Consistent with the possibility of priming, a recent study revealed that the monkey's strategy to solve the DMS task had been based on a simple stimulus–repetition rule (Miller and Desimone 1994). In a revised task to prohibit this strategy, they found enhanced responses when a test stimulus matched a sample stimulus "in mind". Although this enhancement might reflect a recognition process, its relation to stored memory remains unknown.

It should be pointed out that the tuning mechanism and the repetition–suppression effect are fundamentally different phenomena. We compared a single-cell response to different forms (original pattern vs. transforms), whereas other studies compared a single-cell response to the same form with different familiarity. Hence, the transforms that we have used should not be confused with "novel" forms. Furthermore, the effect of tuning mechanism is apparently opposite, because IT cells show significantly stronger responses to learned, or repetitively presented, original patterns. We suspect that the consolidated neuronal responses to learned complex forms are produced and maintained by the tuning mechanism and are thus resistant to the repetition–suppression effect. To acquire and consolidate essential information selectively, it would be useful to adopt stimulus-specific tuning and stimulus-nonspecific suppression mechanisms.

CATEGORIZATION AS EVIDENCE FOR NEURONAL TUNING MECHANISM

The functional significance of the neuronal tuning mechanism is supported by clinical observations that bilateral anterior temporal lesions virtually preclude any visual recognition at the level of unique identity or fine subordinate categorization (Damasio 1990; Damasio et al. 1990). These agnosic patients can recognize neither unique faces nor individual items such as pets, houses, and cars. One key factor for this

across-category deficit is the *similarity* of their forms among exemplars. We propose that the neuronal tuning mechanism in the temporal association area subserves category-related recognition that requires fine discrimination among similar items in each category. It has been reported that monkeys with IT lesions did not show categorical perception of visual stimuli (Iversen and Weiskrantz 1967; Dean 1976; Wilson and DeBauche 1981). Hence, AIT neurons would be critical for categorization of various forms, thereby organizing the stored information by the neuronal tuning mechanism.

The Associative Mechanism in the AIT Cortex

Most of our long-term memories of episodes or objects are organized so that we can retrieve them by association. Anderson and Bower (1980) proposed that human memory only stores "propositions", which are conceived as "structured bundles of associations between elementary ideas or concepts". We can apply this notion to object recognition: Visual memory stores forms as structured bundles of associations between elementary views of objects. The neurophysiological evidence of associative mechanisms in visual memory has been found only recently (Sakai and Miyashita 1993a). We reported previously that a picture-selective delay activity reflected association caused by the fixed order of picture presentation in the DMS task (Miyashita 1988). Evidence from this experiment was restricted to implicit learning, because subjects could solve the task without memorizing the sequence of pictures. In the PA task, explicit associative learning can be imposed to assess the long-term memory more directly.

PAIR-CODING NEURON

In the AIT cortex we found one type of neuron (*pair-coding neuron*), which manifested selective responses to both of paired associates (Sakai and Miyashita 1991). The properties of pair-coding neurons indicate that memory storage is organized such that single neurons can code both paired associates in the PA task (Fig. 6). This result provides new evidence that single neurons acquire form selectivity through associative learning. Here, this type of coding is termed *associative coding*, in which the involvement of associative learning is essential for memory storage.

The associative coding proposed here provides one organizing principle by which the special selectivity of neuronal responses is produced. The spatiotemporal patterns of neuronal discharges selective to object forms thus constitute the basis of ensemble coding. A possible molecular mechanism of the associative coding lies in the change of
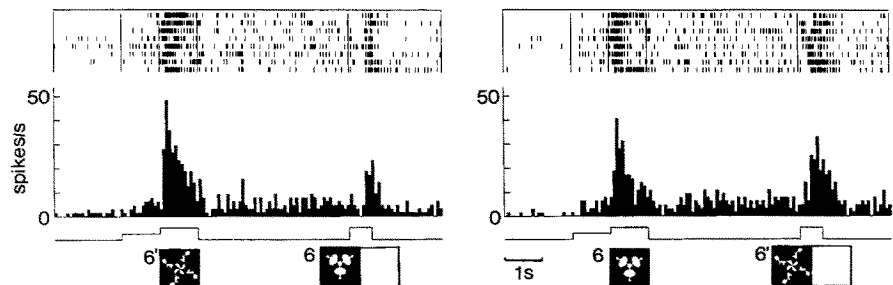

Figure 6: Responses of a pair-coding neuron, which exhibited form-selective activity during the cue period. (*Left*) Trials for cue 6' that elicited the strongest cue response. (*Right*) Trials for cue 6 that elicited the second strongest cue response.

synaptic connections through repetitive learning (Frégnac et al. 1988; Artola and Singer 1990), whereby two inputs are always paired with each other. This mechanism is supported further by a recent model of the neuronal network that learns invariance to shift in retinal position through exposure to temporal sequences of patterns undergoing the transformation (Földiák 1991). The associative mechanism based on temporal contiguity will be discussed further in the section The Implicit Association Process and Object Recognition (below).

PAIR–RECALL NEURON

We found another type of neuron (*pair–recall neuron*) that is presumably involved in the process of memory retrieval (Sakai and Miyashita 1991). Recall that in our experiments the monkey is required to recognize the paired associate of a cue stimulus after a delay period. Pair–recall neurons exhibited form-selective delay activity (Fig. 7). This response is closely coupled with the paired associate that is not actually seen but retrieved by the cue stimulus. There are two possibilities for the critical process during the delay period. One is to hold a *retrospective code*, which is a cue stimulus, in working memory. The other is to generate a *prospective code* by converting a cue into its paired associate. The increasing delay activity of the pair–recall neurons is consistent with the claim that subjects can employ a prospective code. On the grounds that the AIT cortex serves as the memory storehouse (see A Model of the Cognitive Memory System, below), these neurons could serve as memory storage elements, also activated in the retrieval process. The finding of pair–recall neurons is the first neurophysiological demonstration that visual imagery is also implemented by the same neuronal mechanism that subserves memory retrieval.

INTERACTION BETWEEN MEMORY MECHANISMS

Pair-coding neurons and pair–recall neurons can participate, respectively, in the coding and recall processes of visual long-term memory. The pair–recall neuron represents a typical example of interaction between sustained activation and association among the neuronal mechanisms that subserve the formation or expression of memory traces (Desimone 1992). Evidence for form-selective or color-selective sustained activation has been documented in the IT cortex (Fuster and Jervey 1982; Miyashita and Chang 1988; Fuster 1990; Chelazzi et al. 1993) and the
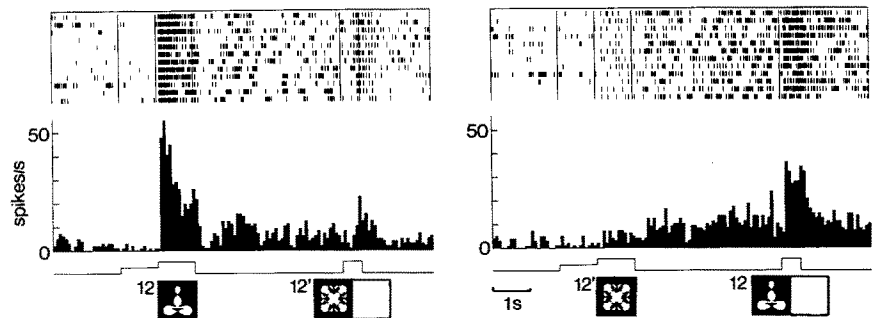
Figure 7: Responses of a pair–recall neuron, which exhibited form-selective activity during the delay period. (*Left*) Trials for cue 12 that elicited the strongest cue response. (*Right*) Trials for cue 12' that elicited the delay response, presumably reflecting retrieval of the paired associate 12. Note the tonic increasing activity during the delay period, which is much stronger than the cue response.

prefrontal cortex (Wilson et al. 1993). Furthermore, there is another type of interaction among neuronal mechanisms. The fact that each pair-coding neuron responds to both paired associates suggests an important role of interaction between neuronal tuning and associative mechanisms. When an AIT neuron responds selectively to a pair of patterns, this interaction predicts that there are at least two separate local maxima in the parameter space.

In the human left temporal neocortex, Haglund et al. (1994) reported increased neuronal activity at memory input and the initial retrieval of an item from recent verbal memory of picture slides. Furthermore, repeated retrievals caused rapid fading of neuronal activity, which is neither a general decrease in activity nor habituation owing to the familiarity of objects presented. This observation suggests interaction between several mechanisms for memory acquisition and retrieval.

CONTRIBUTION OF THE MEDIAL TEMPORAL LOBE TO ASSOCIATIVE MECHANISMS

The medial temporal region has been considered to be essential for the memory consolidation process, by which evanescent information is transformed into an enduring representation for long-term memory (Milner 1972; Squire 1992). In clinical neuropsychology, human memory is assessed by the paired–associate learning test, in which a series of paired words or figures is presented and the subject is then asked to retrieve the other pair member associated with each cue (Wechsler 1987). Patients with lesions of the temporal lobe show marked impairment in this test (Meyer and Yates 1955; Milner 1967; Jones 1974; Petrides 1985; Goldstein et al. 1988). It is also notable that amnesiac patients with bilateral medial temporal lobe damage could not derive benefit from the instructions to use visual imagery in the recall of verbal paired associates (Jones 1974). This evidence suggests the contribution of the medial temporal region in the associative mechanisms. Furthermore, a recent lesion study by Murray et al. (1993) has clearly demonstrated that monkeys with bilateral removal of the medial temporal lobe were severely retarded in relearning the preoperatively trained set of paired associates. The type of memory that this task used would therefore correspond to one that relies on the integrity of these structures.

Form Perception Process from V1 to the AIT Cortex

We now extend the neuronal tuning and associative mechanisms to essential processes in the form perception, thereby providing experimental basis for our model (see Fig. 8, below).

NEURONAL TUNING MECHANISMS IN EARLY SENSORY AREAS

We define neuronal tuning as plastic changes of neuronal response properties in adaptation to a sensory experience. Neuronal tuning includes such changes as those in a neuron's receptive field or its response selectivity. V1 neurons are subject to tuning processes for elementary features of visual stimuli, such as oriented contours, through visual experience in the critical period (Blakemore and Cooper 1970; Hirsch and Spinelli 1970; Stryker et al. 1978; Carlson et al. 1986). Recent studies have revealed that cortical maps for retinotopy, somatotopy, and tonotopy in the primary sensory areas can be extensively reorganized and tuned to manipulated sensory inputs even in adulthood (Merzenich 1987; Kaas 1991; Ramachandran 1993). One of the interesting topics in this line of study is the "filling in" of scotomas caused by lesions of the retina, whereby cortical areas that were initially silenced by the lesions

acquired new receptive fields surrounding those lesions (Kaas et al. 1990; Heinen and Skavenski 1991; Gilbert and Wiesel 1992; Darian-Smith and Gilbert 1994). Furthermore, a psychological technique for creating "artificial scotoma" (Ramachandran and Gregory 1991; Kapadia et al. 1994) has been applied in neurophysiological examination (Pettet and Gilbert 1992; De Weerd et al. 1993). These studies clearly demonstrated that receptive fields in early visual areas can be dynamically altered by the modified sensory context. Similar neuronal tuning mechanisms may also subserve experience-dependent perceptual learning in early visual processing (Karni and Sagi 1993). Furthermore, the neuronal tuning mechanisms found in the temporal association area could be important for perceptual learning in later visual processing, as suggested by a recent study of face-selective neurons (Rolls et al. 1993).

ATTENTIONAL INFLUENCE ON VISUAL AREAS

In primates, form information is subsequently processed in a hierarchy of visual areas that include V1, V2, V4, and the AIT cortex (Van Essen et al. 1992; Merigan and Maunsell 1993). In accord with the intermediate role of V4, some neurons in V4 show selective responses to concentric gratings, which are elementary components of FD patterns (Gallant et al. 1993). There is a claim that the IT region is composed of columnar modules, whose critical features are moderately complex forms, sometimes with color or texture (Fujita et al. 1992; Tanaka 1993). These columnar modules would provide the structural basis for the categorial organization in the AIT cortex, as discussed previously in the section Categorization as Evidence for Neuronal Tuning Mechanisms. It remains to be studied whether these columns found in anesthetized monkeys also function in behaving animals, because attentional gates may exert significant influence on the neuronal activity.

It has been reported that extraretinal factors such as attention modulate the response of IT cells (Moran and Desimone 1985; Richmond and Sato 1987) as well as V4 cells (Haenny et al. 1988; Spitzer et al. 1988), but have little influence on the response of V1 cells (Moran and Desimone 1985; Haenny and Schiller 1988). However, a recent single-unit study clearly demonstrated the neuronal correlates of focal attention in V1 and V2 as well as V4 (Motter 1993; 1994). Attentional influence on early visual areas during perceptual processes has been elucidated further by a computational model proposed by Olshausen et al. (1993). The essential point of their model is that "control neurons" selectively route topographical information from a window of attention in V1 to higher visual areas. They hypothesized that the pulvinar nucleus may function as these control neurons. The pulvinar nucleus is retinotopically organized and reciprocally connected to all visual areas in the occipitotemporal pathway, including V1, V2, V4, and IT (Robinson and Petersen 1992). The role of the pulvinar nucleus in visual attention has been confirmed by positron emission tomography (PET) data (LaBerge and Buchsbaum 1990) and lesion studies (Rafal and Posner 1987).

IS FORM REPRESENTATION 2D VIEWER-CENTERED OR 3D OBJECT-CENTERED?

One of the most fundamental problems in object recognition is how a three-dimensional (3D) object form in the real world is represented in long-term memory. Marr and Nishihara (1978) proposed that a form representation for recognition should use a 3D *object-centered* coordinate system, which is transformed from a 2D *viewer-centered*

coordinate system. In a viewer-centered coordinate system, an object's geometry is specified relative to the viewer. In an object-centered coordinate system, an object's geometry is specified relative to the object itself. The latter emphasizes the computation of a form-specific description that is independent of the vantage point (Biederman 1987). Whereas many face-selective cells in STS respond differentially to different views of a head (viewer-centered) (Perrett et al. 1985), some cells are responsive to multiple views of the head on the basis of face identity (object-centered) (Perrett et al. 1984). There is another report that some STS cells show intermediate responses between viewer-centered and object-centered representations (Hasselmo et al. 1989). However, the evidence of object-centered representation is still not conclusive, because "some cells allocated to this class may have been responding to a simple feature common to all views, such as the presence of hair, rather than to the head itself" (Perrett et al. 1985).

In contrast to the above 3D object-centered model, recent studies propose that multiple 2D views are directly stored for each object representation (Longuet-Higgins 1990; Cavanagh 1993). Ullman and Basri (1989) presented a model in which a 3D object is represented by the linear combination of a small number of 2D views of the object. Poggio and Edelman (1990) extended this idea further to the view approximation model, such that a neuronal network learns to recognize 3D objects if the network has been exposed to enough views of a given object. In this scheme, the network, after being trained on several tens of random views of a wire-frame 3D object, maps any new view of the same object into a standard view (e.g., one of the training views). The results of a psychophysical study (Bülthoff and Edelman 1992) and a behavioral study (Logothetis et al. 1994) support a viewer-centered recognition by interpolating between a small number of stored views. Furthermore, Cavanagh (1991) demonstrated that recognition for shadowed familiar figures is possible even though no 3D representation can be constructed from image data. To establish that memory representations of object forms are multiple 2D views rather than internal 3D models, we need more supporting data that can elucidate the underlying processes for object recognition.

THE IMPLICIT ASSOCIATIVE PROCESS AND OBJECT RECOGNITION

We propose further a possibility that associative mechanisms are shared by two neuronal processes: One is the explicit process that is critical in the PA task, whereas the other is the implicit or automatic process. The latter implicit associative process may subserve object recognition. According to Marr and Nishihara (1978), "any theory for recognition that is based on a viewer-centered representation must treat distinct views of an object essentially as distinct objects." They argued further that, "the important characteristic of this approach is that it requires a potentially large store of descriptions in memory in exchange for a reduction in the magnitude and complexity of the computations that would otherwise be required to compensate for the effects of perspective" (Marr and Nishihara 1978). In contrast, the implicit associative process can treat distinct views of an object as common views of the same object because these views are overwritten in the common neuronal network that consists of many pair-coding neurons. It therefore requires a minimum store of descriptions in memory without any compensation for the effects of perspective. Whereas the identity of an object is guaranteed by its

several views encoded by pair-coding neurons, the independence of each view is still preserved by AIT neurons that respond selectively to each view, as we have already found neurons of both types. Using this scheme, one standard view can be naturally associated with any observed views of the same object, thus enabling correct recognition. The face-selective cells that are responsive to multiple views of the head can also be explained within this scheme, such that the invariance over changes of vantage points was actually acquired through associative learning. As noted by Perrett et al. (1987), "for the majority of objects only learning can specify that the back and front views of an object belong to the same object."

In our visual world, multiple views of an object are nearly always presented in succession, resulting from relative movement between the observer's eyes and the object. This situation is an example of the implicit associative process in that two different views of an object are associated and memorized by AIT neurons (Sakai and Miyashita 1991). As Stryker (1991) pointed out, "if the temporal cortex contains a general mechanism for the association of successively presented stimuli, then these views would automatically become associated, obviating the need for special mechanisms for doing geometrical transformations or for hierarchically associating conjunctions of trigger features." Such an associative mechanism based on temporal contiguity agrees well with the principle of generic image of sampling, proposed by Nakayama and Shimojo (1992). This is a theoretical framework based on the idea of a mobile observer sampling images from random vantage points in space. For example, *generic* or usual images of a cube have three faces; *accidental* or unusual images of a cube have two or only one face. This principle indicates that, "when faced with more than one surface interpretation of an image, the visual system assumes it is viewing the scene from a generic, not an accidental, vantage point" (Nakayama and Shimojo 1992). They argued further that the visual system learns associative connections between images and surfaces through the observer's experience of sampling. Incorporating this significant conceptual framework, the implicit associative process can be elucidated further: The relationship of generic 2D views is automatically acquired by the neuronal mechanism of association.

A Model of the Cognitive Memory System

A model of the cognitive memory system that we have proposed (Sakai and Miyashita 1993a, 1994b) is shown in Figure 8A. This scheme is based on structures and functions of the visual memory system for object recognition (Fig. 8B), and it contains some ideas that have been put forward by several researchers (Gross 1973; Mishkin 1982; Squire and Zola-Morgan 1991; Van Essen et al. 1992; Ungerleider and Haxby 1994).

In the memory acquisition process, sensory information is transformed into a memory code of neuronal responses (encoding) with the bottom-up information flow from feature analyzers to a memory storehouse. The primary visual area and the prestriate area serve as feature analyzers in vision (Van Essen et al. 1992; Ungerleider and Haxby 1994). A possible candidate for the memory storehouse is the temporal association area (Gross 1973; Mishkin 1982; Sakai and Miyashita 1993a). In visual perception, prominent features in a visual field are selected and located by a focal attention controller. As noted above, one candidate for the focal attention controller is the pulvinar nucleus. We hypothesize
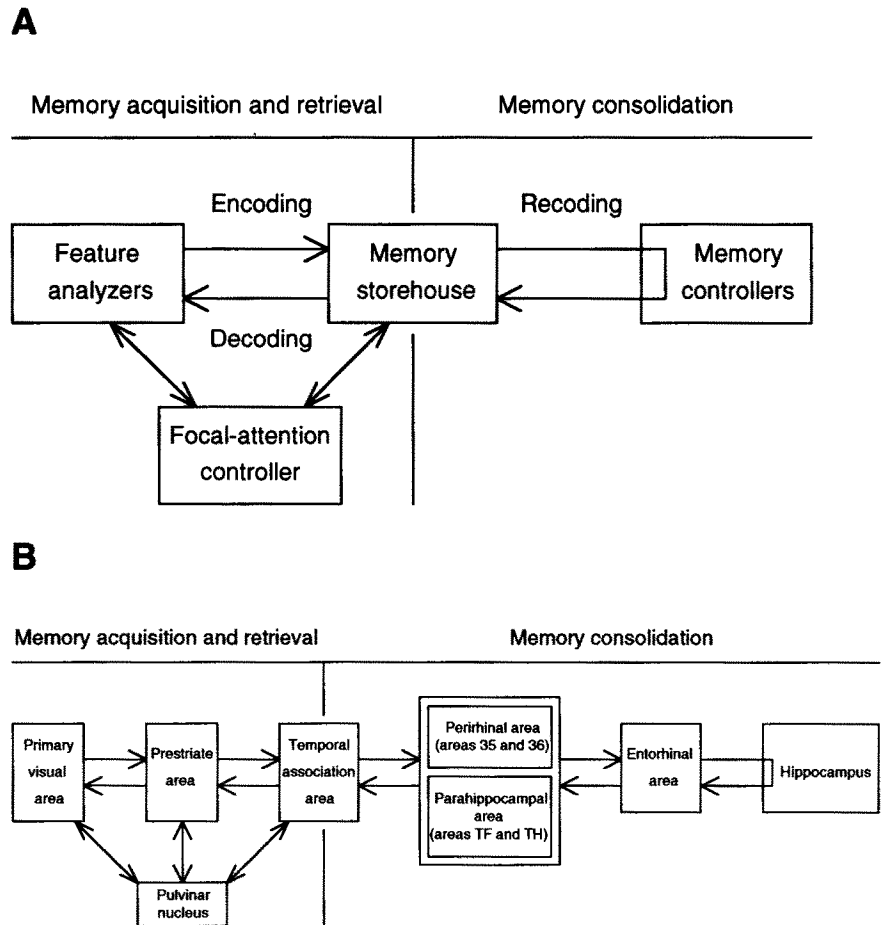
A

Memory acquisition and retrieval   **Memory consolidation**

B

Memory acquisition and retrieval   **Memory consolidation**

Figure 8:   (*A*) A model of the cognitive memory system that unifies perception and imagery. Perception is implemented by the interaction between memory acquisition (encoding) and focal attention mechanisms. Imagery is implemented by the interaction between memory retrieval (decoding) and focal attention mechanisms. (See text for further explanation.) (*B*) Structures of the visual memory system for object recognition. Information flow between these structures corresponds to information flow indicated in *A*. Brodmann's areas 35 and 36, and von Economo's areas TF and TH are indicated. Note that most of the pathways between cortical visual areas are anatomically reciprocal, supporting their functional roles in encoding and decoding processes.

that perception is implemented by the interaction between memory acquisition (encoding) and focal attention mechanisms (Fig. 8A).

To establish a long-lasting representation in the memory storehouse, memory controllers are responsible for consolidation. During this memory consolidation process, the memory code is reorganized dynamically (recoding) by the interaction between the memory storehouse and memory controllers. Medial temporal structures, including the hippocampus, are regarded as memory controllers because of their plasticity (e.g., long-term potentiation and depression) (Bliss and Collingridge 1993; Malenka 1993), neural networks (e.g., autoassociation) (McNaughton and Morris 1987; Rolls 1989), clinical observations (e.g., medial temporal lobe amnesia) (Milner 1972; Squire

1992), and lesion studies (Squire and Zola-Morgan 1991; Mishkin and Murray 1994).

In the memory retrieval process, memory codes stored in the memory storehouse are decomposed into more elementary attributes (decoding) by feature analyzers through a back-projection pathway. We hypothesize that imagery is implemented by the interaction between memory retrieval (decoding) and focal attention mechanisms. As Kosslyn (1988) pointed out, "imagery consists of brain states like those that arise during perception but occurs in the absence of the appropriate immediate sensory input". The parallel mechanisms for imagery and attention have been supported by psychological evidence (Farah 1989). According to our scheme, visual imagery is generated by top-down activation of perceptual representations that are selected and located by the focal–attention controller. It is also possible that a signal from the focal–attention controller gates the top-down information flow toward early visual areas that function in the decoding process. This model predicts that the extent of visual areas devoted to decoding mental images is controlled dynamically by focal attention. The supporting evidence for this proposal on visual imagery has been discussed elsewhere (Sakai and Miyashita 1994b).

Conclusions

We analyzed the fine-form selectivity of AIT neurons and found memory mechanisms of neuronal tuning and association that determine learning-dependent selectivity for object forms. The acquired form selectivity is a key feature in the capacity of temporal cortical neurons to establish the form representation with multiple 2D views. On the grounds that the temporal association area is a memory storehouse in object vision, the neuronal mechanisms reported here could be useful in the acquisition and retrieval of visual long-term memory.

Acknowledgments

References

Adelson, E.H. 1993. Perceptual organization and the judgment of brightness. Science 262: 2042–2044.

Albright, T.D., R.A. Charles, and C.G. Gross. 1985. Inferior temporal neurons do not seem to code shape by the method of fourier descriptors. Soc. Neurosci. Abstr. 11: 1013.

Anderson, J.R. and G.H. Bower. 1980. Human associative memory: A brief edition. Lawrence Erlbaum, Hillsdale, NJ.

Artola, A. and W. Singer. 1990. The involvement of N-methyl-D-aspartate receptors in induction and maintenance of long-term potentiation in rat visual cortex. Eur. J. Neurosci. 2: 254–269.

Baylis, G.C. and E.T. Rolls. 1987. Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. Exp. Brain Res. 65: 614–622.

Biederman, I. 1987. Recognition-by-components: A theory of human image understanding. Psychol. Rev. 94: 115–147.

Blakemore, C. and G.F. Cooper. 1970. Development of the brain depends on the visual environment. Nature 228: 477–478.

Blasdel, G.G. 1992. Orientation selectivity, preference, and continuity in monkey striate cortex. *J. Neurosci.* **12:** 3139–3161.

Bliss, T.V.P. and G.L. Collingridge. 1993. A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature* **361:** 31–39.

Bonhoeffer, T. and A. Grinvald. 1991. Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature* **353:** 429–431.

Bülthoff, H.H. and S. Edelman. 1992. Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Natl. Acad. Sci.* **89:** 60–64.

Carlson, M., D.H. Hubel, and T.N. Wiesel. 1986. Effects of monocular exposure to oriented lines on monkey striate cortex. *Dev. Brain Res.* **25:** 71–81.

Cavanagh, P. 1991. What's up in top-down processing? In *Representations of vision: Trends and tacit assumptions in vision research* (ed. A. Gorea), pp. 295–304. Cambridge University Press, Cambridge, U.K.

———. 1993. The perception of form and motion. *Curr. Opin. Neurobiol.* **3:** 177–182.

Cave, C.B. and S.M. Kosslyn. 1993. The role of parts and spatial relations in object identification. *Perception* **22:** 229–248.

Chelazzi, L., E.K. Miller, J. Duncan, and R. Desimone. 1993. A neural basis for visual search in inferior temporal cortex. *Nature* **363:** 345–347.

Damasio, A.R. 1990. Category-related recognition defects as a clue to the neural substrates of knowledge. *Trends Neurosci.* **13:** 95–98.

Damasio, A.R., H. Damasio, and D. Tranel. 1990. Impairments of visual recognition as clues to the processes of memory. In *Signal & sense: Local and global order in perceptual maps* (ed. G.M. Edelman, W.E. Gall, and W.M. Cowan), pp. 451–473. Wiley, New York.

Darian-Smith, C. and C.D. Gilbert. 1994. Axonal sprouting accompanies functional reorganization in adult cat striate cortex. *Nature* **368:** 737–740.

De Weerd, P., R. Gattass, R. Desimone, and L.G. Ungerleider. 1993. Center-surround interactions in areas V2/V3: A possible mechanism for filling-in? *Soc. Neurosci. Abstr.* **19:** 27.

Dean, P. 1976. Effects of inferotemporal lesions on the behavior of monkeys. *Psychol. Bull.* **83:** 41–71.

Desimone, R. 1992. The physiology of memory: Recordings of things past. *Science* **258:** 245–246.

Desimone, R., T.D. Albright, C.G. Gross, and C. Bruce. 1984. Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* **4:** 2051–2062.

Fahy, F.L., I.P. Riches, and M.W. Brown. 1993. Neuronal activity related to visual recognition memory: Long-term memory and the encoding of recency and familiarity information in the primate anterior and medial inferior temporal and rhinal cortex. *Exp. Brain Res.* **96:** 457–472.

Farah, M.J. 1989. Mechanisms of imagery-perception interaction. *J. Exp. Psychol. Hum. Percept. Perform.* **15:** 203–211.

Földiák P. 1991. Learning invariance from transformation sequences. *Neural Comp.* **3:** 194–200.

Frégnac, Y., D. Shulz, S. Thorpe, and E. Bienenstock. 1988. A cellular analogue of visual cortical plasticity. *Nature* **333:** 367–370.

Fujita, I., K. Tanaka, M. Ito, and K. Cheng. 1992. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* **360:** 343–346.

Fuster, J.M. 1990. Inferotemporal units in selective visual attention and short-term memory. *J. Neurophysiol.* **64:** 681–697.

Fuster, J.M. and J.P. Jervey. 1982. Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. *J. Neurosci.* **2:** 361–375.

Gaffan, D. 1985. Hippocampus: Memory, habit and voluntary movement. *Phil. Trans. R. Soc. Lond. B.* **308:** 87–99.

Gallant, J.L., J. Braun, and D.C. Van Essen. 1993. Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex. *Science* **259:** 100–103.

Gilbert, C.D. and T.N. Wiesel. 1992. Receptive field dynamics in adult primary visual cortex. *Nature* **356:** 150–152.

Goldstein, L.H., A.G.M. Canavan, and C.E. Polkey. 1988. Verbal and abstract designs paired associate learning after unilateral temporal lobectomy. *Cortex* **24:** 41–52.

Gross, C.G. 1973. Visual functions of inferotemporal cortex. In *Handbook of sensory physiology, vol. VII/3, Central processing of visual information, part B* (ed. R. Jung), pp. 451–482. Springer-Verlag, Berlin, Germany.

———. 1992. Representation of visual stimuli in inferior temporal cortex. *Phil. Trans. R. Soc. Lond. B.* **335:** 3–10.

Gross, C.G., C.E. Rocha-Miranda, and D.B. Bender. 1972. Visual properties of neurons in inferotemporal cortex of the macaque. *J. Neurophysiol.* **35:** 96–111.

Haenny, P.E. and P.H. Schiller. 1988. State dependent activity in monkey visual cortex, I. Single cell activity in V1 and V4 on visual tasks. *Exp. Brain Res.* **69:** 225–244.

Haenny, P.E., J.H.R. Maunsell, and P.H. Schiller. 1988. State dependent activity in monkey visual cortex, II. Retinal and extraretinal factors in V4. *Exp. Brain Res.* **69:** 245–259.

Haglund, M.M., G.A. Ojemann, T.W. Schwartz, and E. Lettich. 1994. Neuronal activity in human lateral temporal cortex during serial retrieval from short-term memory. *J. Neurosci.* **14:** 1507–1515.

Hasselmo, M.E., E.T. Rolls, G.C. Baylis, and V. Nalwa. 1989. Object-centered encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. *Exp. Brain Res.* **75:** 417–429.

Heinen, S.J. and A.A. Skavenski. 1991. Recovery of visual responses in foveal V1 neurons following bilateral foveal lesions in adult monkey. *Exp. Brain Res.* **83:** 670–674.

Hirsch, H.V.B. and D.N. Spinelli. 1970. Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats. *Science* **168:** 869–871.

Hubel, D.H. and T.N. Wiesel. 1959. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* **148:** 574–591.

Iversen, S. and L. Weiskrantz. 1967. Perception of redundant cues by monkeys with inferotemporal lesions. *Nature* **214:** 241–243.

Jones, M.K. 1974. Imagery as a mnemonic aid after left temporal lobectomy: contrast between material-specific and generalized memory disorders. *Neuropsychologia* **12:** 21–30.

Kaas, J.H. 1991. Plasticity of sensory and motor maps in adult mammals. *Annu. Rev. Neurosci.* **14:** 137–167.

Kaas, J.H., L.A. Krubitzer, Y.M. Chino, A.L. Langston, E.H. Polley, and N. Blair. 1990. Reorganization of retinotopic cortical maps in adult mammals after lesions of the retina. *Science* **248:** 229–231.

Kapadia, M.K., C.D. Gilbert, and G. Westheimer. 1994. A quantitative measure for short-term cortical plasticity in human vision. *J. Neurosci.* **14:** 451–457.

Karni, A. and D. Sagi. 1993. The time course of learning a visual skill. *Nature* **365:** 250–252.

Kosslyn, S.M. 1988. Aspects of cognitive neuroscience of mental imagery. *Science* **240:** 1621–1626.

LaBerge, D. and M.S. Buchsbaum. 1990. Positron emission tomographic measurements of pulvinar activity during an attention task. *J. Neurosci.* **10:** 613–619.

Li, L., E.K. Miller, and R. Desimone. 1993. The representation of stimulus familiarity in anterior inferior temporal cortex. *J. Neurophysiol.* **69:** 1918–1929.

Logothetis, N.K., J. Pauls, H.H. Bülthoff, and T. Poggio. 1994. View-dependent object recognition by monkeys. *Curr. Biol.* **4:** 401–414.

Longuet-Higgins, H.C. 1990. Recognizing three dimensions. *Nature* **343:** 214–215.

Malenka, R.C. 1993. Long-term depression: Not so depressing after all. *Proc. Natl. Acad. Sci.* **90:** 3121–3123.

Marr, D. 1982. *Vision*. Freeman, New York.

Marr, D. and H.K. Nishihara. 1978. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond. B.* **200:** 269–294.

McNaughton, B.L. and R.G.M. Morris. 1987. Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci.* **10:** 408–415.

Merigan, W.H. and J.H.R. Maunsell. 1993. How parallel are the primate visual pathways? *Annu. Rev. Neurosci.* **16:** 369–402.

Merzenich, M.M. 1987. Dynamic neocortical processes and the origins of higher brain functions. In *The neural and molecular bases of learning* (ed. J.-P. Changeux and M. Konishi), pp. 337–358. John Wiley & Sons, Chichester, U.K.

Meyer, V. and A.J. Yates. 1955. Intellectual changes following temporal lobectomy for psychomotor epilepsy: Preliminary communication. *J. Neurol. Neurosurg. Psychiatry* **18:** 44–52.

Miller, E.K. and R. Desimone. 1994. Parallel neuronal mechanisms for short-term memory. *Science* **263:** 520–522.

Miller, E.K., P.M. Gochin, and C.G. Gross. 1991a. Habituation-like decrease in the responses of neurons in inferior temporal cortex of the macaque. *Vis. Neurosci.* **7:** 357–362.

Miller, E.K., L. Li, and R. Desimone. 1991b. A neural mechanism for working and recognition memory in inferior temporal cortex. *Science* **254:** 1377–1379.

Milner, B. 1967. Brain mechanisms suggested by studies of temporal lobes. In *Brain mechanisms underlying speech and language* (ed. C.H. Millikan and F.L. Darley), pp. 122–145. Grune & Stratton, New York.

———. 1972. Disorders of learning and memory after temporal lobe lesions in man. *Clin. Neurosurg.* **19:** 421–446.

Mishkin, M. 1982. A memory system in the monkey. *Phil. Trans. R. Soc. Lond. B.* **298:** 85–95.

Mishkin, M. and E.A. Murray. 1994. Stimulus recognition. *Curr. Opin. Neurobiol.* **4:** 200–206.

Miyashita, Y. 1988. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* **335:** 817–820.

Miyashita, Y. and H.S. Chang. 1988. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature* **331:** 68–70.

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

Motter, B.C. 1993. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.* **70:** 909–919.

———. 1994. Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J. Neurosci.* **14:** 2178–2189.

Murray, E.A., D. Gaffan, and M. Mishkin. 1993. Neural substrates of visual stimulus-stimulus association in rhesus monkeys. *J. Neurosci.* **13:** 4549–4561.

Nakayama, K. and S. Shimojo. 1992. Experiencing and perceiving visual surfaces. *Science* **257:** 1357–1363.

Olshausen, B.A., C.H. Anderson, and D.C. Van Essen. 1993. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J. Neurosci.* **13:** 4700–4719.

Pentland, A.P. 1986. Perceptual organization and the representation of natural form. *Artif. Intell.* **28:** 293–331.

Perrett, D.I., P.A.J. Smith, D.D. Potter, A.J. Mistlin, A.S. Head, A.D. Milner, and M.A. Jeeves. 1984. Neurons responsive to faces in the temporal cortex: Studies of functional organization, sensitivity to identity and relation to perception. *Hum. Neurobiol.* **3:** 197–208.

———. 1985. Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc. R. Soc. Lond. B.* **223:** 293–317.

Perrett, D.I., A.J. Mistlin, and A.J. Chitty. 1987. Visual neurones responsive to faces. *Trends Neurosci.* **10:** 358–364.

Petrides, M. 1985. Deficits on conditional associative-learning tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia* **23:** 601–614.

Pettet, M.W. and C.D. Gilbert. 1992. Dynamic changes in receptive-field size in cat primary visual cortex. *Proc. Natl. Acad. Sci.* **89:** 8366–8370.

Poggio, T. and S. Edelman. 1990. A network that learns to recognize three-dimensional objects. *Nature* **343:** 263–266.

Rafal, R.D. and M.I. Posner. 1987. Deficits in human visual spatial attention following thalamic lesions. *Proc. Natl. Acad. Sci.* **84:** 7349–7353.

Ramachandran, V.S. 1993. Behavioral and magnetoencephalographic correlates of plasticity in the adult human brain. *Proc. Natl. Acad. Sci.* **90:** 10413–10420.

Ramachandran, V.S. and R.L. Gregory. 1991. Perceptual filling in of artificially induced scotomas in human vision. *Nature* **350:** 699–702.

Riches, I.P., F.A.W. Wilson, and M.W. Brown. 1991. The effects of visual stimulation and

memory on neurons of the hippocampal formation and the neighboring parahippocampal gyrus and inferior temporal cortex of the primate. *J. Neurosci.* **11:** 1763–1779.

Richmond, B.J. and T. Sato. 1987. Enhancement of inferior temporal neurons during visual discrimination. *J. Neurophysiol.* **58:** 1292–1306.

Richmond, B.J., L.M. Optican, M. Podell, and H. Spitzer. 1987. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. *J. Neurophysiol.* **57:** 132–146.

Robinson, D.L. and S.E. Petersen. 1992. The pulvinar and visual salience. *Trends Neurosci.* **15:** 127–132.

Rolls, E.T. 1989. Parallel distributed processing in the brain: Implications of the functional architecture of neuronal networks in the hippocampus. In *Parallel distributed processing: Implications for psychology and neurobiology.* (ed. R.G.M. Morris), pp. 286–308. Clarendon Press, Oxford, UK.

Rolls, E.T. and G.C. Baylis. 1986. Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Exp. Brain Res.* **65:** 38–48.

Rolls, E.T., G.C. Baylis, M.E. Hasselmo, and V. Nalwa. 1989. The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey. *Exp. Brain Res.* **76:** 153–164.

Rolls, E.T., M.J. Tovee, and V.S. Ramachandran. 1993. Visual learning reflected in the responses of neurons in the temporal visual cortex of the macaque. *Soc. Neurosci. Abstr.* **19:** 28.

Sakai, K. and Y. Miyashita. 1991. Neural organization for the long-term memory of paired associates. *Nature* **354:** 152–155.

———. 1993a. Memory and imagery in the temporal lobe. *Curr. Opin. Neurobiol.* **3:** 166–170.

———. 1993b. Neuronal tuning to learned complex forms in the primate anterior inferotemporal cortex. *Soc. Neurosci. Abstr.* **19:** 28.

———. 1994a. Neuronal tuning to learned complex forms in vision. *NeuroReport* **5:** 829–832.

———. 1994b. Visual imagery: An interaction between memory retrieval and focal attention. *Trends Neurosci.* **17:** 287–289.

Sáry, G., R. Vogels, and G.A. Orban. 1993. Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science* **260:** 995–997.

Schwartz, E.L., R. Desimone, T.D. Albright, and C.G. Gross. 1983. Shape recognition and inferior temporal neurons. *Proc. Natl. Acad. Sci.* **80:** 5776–5778.

Snedecor, G.W. and W.G. Cochran. 1989. *Statistical methods,* 8th ed. Iowa State University Press, Ames, IA.

Spitzer, H., R. Desimone, and J. Moran. 1988. Increased attention enhances both behavioral and neuronal performance. *Science* **240:** 338–340.

Squire, L.R. 1992. Memory and the Hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* **99:** 195–231.

Squire, L.R. and S. Zola-Morgan. 1991. The medial temporal lobe memory system. *Science* **253:** 1380–1386.

Stryker, M.P. 1991. Temporal associations. *Nature* **354:** 108–109.

Stryker, M.P., H. Sherk, A.G. Leventhal, and H.V.B. Hirsch. 1978. Physiological consequences for the cat's visual cortex of effectively restricting early visual experience with oriented contours. *J. Neurophysiol.* **41:** 896–909.

Tanaka, K. 1993. Neuronal mechanisms of object recognition. *Science* **262:** 685–688.

Tanaka, K., H. Saito, Y. Fukada, and M. Moriya. 1991. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol.* **66:** 170–189.

Tulving, E. and D.L. Schacter. 1990. Priming and human memory systems. *Science* **247:** 301–306.

Ullman, S. and R. Basri. 1989. Recognition by linear combinations of models. In *Artificial intelligence laboratory memo no. 1152.* MIT Artificial Intelligence Laboratory, Cambridge, M.A.

Ungerleider, L.G. and J.V. Haxby. 1994. 'What' and 'where' in the human brain. *Curr. Opin. Neurobiol.* **4:** 157–165.

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

Wechsler, D. 1987. *Wechsler Memory Scale—Revised.* The Psychological Corporation, San Antonio, TX.

Wilson, M. and B.A. DeBauche. 1981. Inferotemporal cortex and categorical perception of visual stimuli by monkeys. *Neuropsychologia* **19:** 29–41.

Wilson, F.A.W., S.P.O. Scalaidhe, and P.S. Goldman-Rakic. 1993. Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* **260:** 1955–1958.

Zahn, C.T. and R.Z. Roskies. 1972. Fourier descriptors for plane closed curves. *IEEE Trans. Comput.* **c-21:** 269–281.